

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 11:29:56 ; Search time 463 seconds

(without alignments)  
439.155 Million cell updates/sec

Title: US-09-702-216-2

Perfect score: 6851

Sequence: 1 MYRKQKPLPLNVASEGEGQI.....IQRGLHRNNAQVENCKPKK 1281

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :			
A_Geneseq_19Jun03.*			
1:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*		
2:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*		
3:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*		
4:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*		
5:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*		
6:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*		
7:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*		
8:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*		
9:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*		
10:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*		
11:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*		
12:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*		
13:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*		
14:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*		
15:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*		
16:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*		
17:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*		
18:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*		
19:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*		
20:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*		
21:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*		
22:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*		
23:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*		
24:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6851	100.0	1281	22	AAE02189 Human breast cancer
2	6851	100.0	1281	24	ABR47615 Breast cancer asso
3	6851	100.0	1294	23	ABJ05569 Breast cancer-asso
4	6843	99.9	1281	21	AAE00190 Breast cancer proc
5	6832	99.7	1294	21	AAAB0550 Human ORF314
6	4212	61.5	785	22	AAAB92512 Human protein sequ
7	286	4.2	1213	22	AAAB41128 Human polypeptide
8	271	4.0	758	22	ABB64656 Drosophila melanog
9	269	3.9	499	24	ABU11707 Human MDR polypep

10	269	3.9	519	21	AAAB2333 Human ORF2097
11	267.5	3.9	1224	22	RAM39342 Human polypeptide
12	256	3.7	734	22	ABB71878 Drosophila melanog
13	256	3.7	1311	22	AAW93979 Human stomach can
14	256	3.7	1311	22	AAAB95671 Human protein sequ
15	251.5	3.7	533	21	AAAY8745 Murine embryonic b
16	248.5	3.6	516	19	AAW70971 Human Ikaros isofo
17	247.5	3.6	950	22	ABB71271 Drosophila melanog
18	246.5	3.6	537	23	ABB79561 Mouse Daedalus pro
19	241	3.5	537	15	AAAR46964 Peptide with Ikar
20	240	3.5	419	23	ABP69811 Human polypeptide
21	239.5	3.5	1296	23	ABG61925 Prostate cancer-as
22	238	3.5	461	17	AAAR92015 Human Ikaros prote
23	238	3.5	461	19	AAW72672 Human Ikaros. Hom
24	238	3.5	461	23	AAO18605 Human Ikaros prote
25	237.5	3.5	519	23	AAAC18606 Murine Iki protein
26	237	3.5	461	19	AAW70964 Human Ikaros isofo
27	235.5	3.4	1181	22	AAAB95553 Murine protein sequ
28	233.5	3.4	518	17	AAAR92017 Murine Ikaros prot
29	233.5	3.4	518	19	AAW72674 Mouse Ikaros mfk-1
30	233.5	3.4	518	19	AAW79966 Mouse Ikaros isofo
31	231	3.4	3572	23	ABG95659 Human. nucleic acid
32	230.5	3.4	568	15	AAAR46965 Ikaros zinc finger
33	228	3.3	351	23	AAU96738 Human cytokine rec
34	225	3.3	334	17	AAAR92020 Ikaros protein. N
35	225	3.3	334	19	AAW72677 Ikaros isoform. U
36	225	3.3	334	19	AAW70969 Human Ikaros prote
37	224.5	3.3	432	17	AAAR92016 Murine Ikaros prot
38	224.5	3.3	432	19	AAW72673 Mouse Ikaros mfk-3
39	224.5	3.3	432	19	AAW70965 Mouse Ikaros isofo
40	223.5	3.3	1696	22	ABG10886 Novel human diagno
41	222	3.2	323	22	AAU15994 Human novel secret
42	222	3.2	323	22	AAU16142 Human novel secret
43	222	3.2	323	24	ABU55063 Human novel polype
44	222	3.2	323	24	ABU55211 Human novel polype
45	220	3.2	757	21	AAAB40777 Human ORF541

ALIGNMENTS

RESULT 1	
AAEC2189	
ID	.AAE02189 standard; Protein; 1281 AA.
XX	AAE02189;
XX	
DT	06-AUG-2001 (first entry)
XX	
DE	Human breast cancer specific gene-3 (BCSG-3) protein.
XX	
KW	Human; breast cancer specific gene-3; BCSG-3; cytostatic; vaccine;
KW	breast cancer; therapeutic; gene therapy.
XX	
OS	Homo sapiens.
XX	
EH	Key Location/Qualifiers
FT	Misc-difference 426 /note= "Encoded by AA"
FT	Misc-difference 1100 /note= "Encoded by A"
XX	
PN	WC200137779-A2.
XX	
PD	31-MAY-2001.
XX	
PF	22-NOV-2000; 2000WQ-US32056.
XX	
PR	23-NOV-1999; 99US-0166973.
XX	
PA	(DIAD-) CIADEXUS INC.
XX	
PI	Salceda S, Caferkey R, Recipon H, Sun Y;

XX WPI: 2001-367602/38.  
 DR N-PSDB; AAD06861.  
 XX  
 PT Novel breast cancer specific gene for diagnosing, monitoring, staging,  
 PT imaging, preventing and treating cancers, particularly breast cancer.  
 XX  
 PS Claim 2; Page 61-66; 66pp; English.  
 XX  
 CC The invention relates human breast cancer specific genes (BCSG's) and  
 CC their corresponding proteins. BCSG is useful for diagnosing, staging,  
 CC monitoring, imaging, preventing and treating breast cancers. BCSG is also  
 CC useful for inducing an immune response against a target cell expressing  
 CC BCSG. The invention also provide methods for detecting genetic lesions or  
 CC mutations in BCSG, thereby determining if a human with the genetic lesion  
 CC is at risk for breast cancer or has breast cancer. BCSG antibodies  
 CC labelled with paramagnetic ions or radioisotopes is useful for imaging  
 CC breast cancers, while BCSG antibodies conjugated to a cytotoxic agent is  
 CC useful for treating breast cancer. BCSG is useful in the rational design  
 CC of new therapeutics for imaging and treating cancers. BCSG is also used  
 CC in gene therapy. The present sequence is human breast cancer specific  
 CC gene-3 (BCSG-3) protein.  
 XX  
 SQ Sequence 1281 AA;

Query Match 100.0%; Score 6851; DB 22; Length 1281;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVRKKNPPLRNVASEGQILEPITGESKVGKNKEFSADQMSENTDQSDAEALNHKEH 60  
 DB 1 MVRKKNPPLRNVASEGQILEPITGESKVGKNKEFSADQMSENTDQSDAEALNHKEH 60  
 QY 61 SLHVQDPSSSKKDLKSAVLSEKAGFNYESPSKGNFPSPFPHDEVTDNRKLAFFSPAAG 120  
 DB 61 SLHVQDPSSSKKDLKSAVLSEKAGFNYESPSKGNFPSPFPHDEVTDNRKLAFFSPAAG 120  
 QY 121 VCEPLKSPQAEADDDQDMACITFSGDSLETKEDQKMSPKATEGTGAQSQANCOGLSPV 180  
 DB 121 VCEPLKSPQAEADDDQDMACITFSGDSLETKEDQKMSPKATEGTGAQSQANCOGLSPV 180  
 QY 181 SVASKNPQVPSDGGVRLNKSKTLLVNDNPDAPLSPETQDFKCNCGYGYGNDPTDJI 240  
 DB 181 SVASKNPQVPSDGGVRLNKSKTLLVNDNPDAPLSPETQDFKCNCGYGYGNDPTDJI 240  
 QY 241 KHPKRYHLGLHNRTRQDAELDSKILALHNMVQFSKDFQKNRSVFSGLQDINSRPV 300  
 DB 241 KHPKRYHLGLHNRTRQDAELDSKILALHNMVQFSKDFQKNRSVFSGLQDINSRPV 300  
 QY 301 LLNGTYDVQVTSGGTIGIGRKTPTDCQNTKYPKCFKCNFTYMGNSSTELCHFLQTHN 360  
 DB 301 LLNGTYDVQVTSGGTIGIGRKTPTDCQNTKYPKCFKCNFTYMGNSSTELCHFLQTHN 360  
 QY 361 KIKASLPSEVAKPSEKSNKSIKALQSSDSGLGKWCQKXITVAKAGDTPVGVSVPIKPL 420  
 DB 361 KIKASLPSEVAKPSEKSNKSIKALQSSDSGLGKWCQKXITVAKAGDTPVGVSVPIKPL 420  
 QY 421 DSSRQNGTEATSYVCKFCFSCESSSLKLEHYGKHQGVQSGGKINPELNDKLSRGSV 480  
 DB 421 DSSRQNGTEATSYVCKFCFSCESSSLKLEHYGKHQGVQSGGKINPELNDKLSRGSV 480  
 QY 481 INQNDLAKSSEGETMTKTDKSSGAKKDFSSKGAEDNMVTSYNCQDFRYSKSHGPV 540  
 DB 481 INQNDLAKSSEGETMTKTDKSSGAKKDFSSKGAEDNMVTSYNCQDFRYSKSHGPV 540  
 QY 541 IVVGPLLRHYQQLHNTKCTIKKCPFCRGLCSPEKHLGEITYPPFACRKNCSHCAALLL 600  
 DB 541 IVVGPLLRHYQQLHNTKCTIKKCPFCRGLCSPEKHLGEITYPPFACRKNCSHCAALLL 600  
 QY 601 HLSFGAAGSRVRKHQCHQCSFTTTPDQVLLFHYESVHESQASDVQKQEAHLQSGDQCSV 660  
 DB 601 HLSFGAAGSRVRKHQCHQCSFTTTPDQVLLFHYESVHESQASDVQKQEAHLQSGDQCSV 660

QY 661 KESKESCTKCDFITOVEEEISRHYRAHSCYKGCROCSFTAAADTCSLLEHENTVHCQEQD 720  
 DB 661 KESKESCTKCDFITOVEEEISRHYRAHSCYKGCROCSFTAAADTCSLLEHENTVHCQEQD 720  
 QY 721 ITTANGEDGHAISTIKKEPKIDFRVYNLLTPDSKXGPEVSESVVKKLEKDGLEKGV 780  
 DB 721 ITTANGEDGHAISTIKKEPKIDFRVYNLLTPDSKXGPEVSESVVKKLEKDGLEKGV 780  
 QY 781 WTSSDDLKRVTRWGADILRGSPSYTOASLGLLTPVSGTQEQTKTLRDSNPVAAHLAR 840  
 DB 781 WTSSDDLKRVTRWGADILRGSPSYTOASLGLLTPVSGTQEQTKTLRDSNPVAAHLAR 840  
 QY 841 PIYGLAVETKCFLOGAPAGGKSGALPOQVPASGENSKDESQSLRRRSGGVFCANCL 900  
 DB 841 PIYGLAVETKCFLOGAPAGGKSGALPOQVPASGENSKDESQSLRRRSGGVFCANCL 900  
 QY 901 TTKTSLMKPNANGYVVCNACGLYQKLHSTPRPLNIIKQNGEQIIRRTKRLNPEALQA 960  
 DB 901 TTKTSLMKPNANGYVVCNACGLYQKLHSTPRPLNIIKQNGEQIIRRTKRLNPEALQA 960  
 QY 961 EQLNKQORGSENEOVNGSPLERRSEDLHTESHQREIPLPSLSKYEAQGSILTKSHSAQOPV 1020  
 DB 961 EQLNKQORGSENEOVNGSPLERRSEDLHTESHQREIPLPSLSKYEAQGSILTKSHSAQOPV 1020  
 QY 1021 LVSQTLDIHKEMQPLHIQIKSPQESTGDPGNSSSVSEKSGSSERGSPIEKYMRPAKHNY 1080  
 DB 1021 LVSQTLDIHKEMQPLHIQIKSPQESTGDPGNSSSVSEKSGSSERGSPIEKYMRPAKHNY 1080  
 QY 1081 SPPGSPISKYQYPLFGPLPFVHNDPQSEADMLRFWSKYKLSVPGNPHYLSHVPGLPNQCQ 1140  
 DB 1081 SPPGSPISKYQYPLFGPLPFVHNDPQSEADMLRFWSKYKLSVPGNPHYLSHVPGLPNQCQ 1140  
 QY 1141 YVPYPTNLPHPSAVGSDNDIPDLAIKHSRPGPTANGASKETKAPPVYNSGPNLVV 1200  
 DB 1141 YVPYPTNLPHPSAVGSDNDIPDLAIKHSRPGPTANGASKETKAPPVYNSGPNLVV 1200  
 QY 1201 KTEKVDSTQDELSCTKVHCGIVFDEVVYALHVSCHGDSGFFQCS:COHLCTDKYDFTT 1260  
 DB 1201 KTEKVDSTQDELSCTKVHCGIVFDEVVYALHVSCHGDSGFFQCS:COHLCTDKYDFTT 1260  
 QY 1261 HIQRLHNNACVCKGPKKE 1281  
 DB 1261 HIQRLHNNACVCKGPKKE 1281

RESULT 2  
 ABR47615  
 ID ABR47615 standard; Protein; 1281 AA.  
 XX ABR47615;  
 XX 12-JUN-2003 (first entry)  
 XX Breast cancer associated protein sequence SEQ ID NO: 472.  
 XX Human; breast cancer; cytostatic; gene therapy.  
 XX Homo sapiens.  
 XX W02003004989-A2.  
 XX 16-JAN-2003.  
 XX 21-JUN-2002; 2002WO-US-9669.  
 XX 21-JUN-2001; 2001US-299887P.  
 XX 27-JUN-2001; 2001US-301572P.  
 XX 18-JUL-2001; 2001US-308501P.  
 XX 25-SEP-2001; 2001US-325002P.  
 XX 05-MAR-2002; 2002US-362585P.  
 XX 14-MAY-2002; 2002US-390391P.  
 XX (MILL-) MILLENIUM PHARM INC.











QY 1081 SPSPSPKIEKYQYPLGFLPFVHNDFOSEADMLRFMSKYKLSVPGNPHYLSHVPGLPNQCQ 1140  
 Db 1094 SPSPSPKIEKYQYPLGFLPFVHNDFOSEADMLRFMSKYKLSVPGNPHYLSHVPGLPNQCQ 1153  
 QY 1141 YVPYPTNLPHFSAVGSNDIPDLDAIKHSRPGPTANGASKEKTKAPPNVKNEGPLNVV 1200  
 Db 1154 YVPYPTNLPHFSAVGSNDIPDLDAIKHSRPGPTANGASKEKTKAPPNVKNEGPLNVV 1213  
 QY 1201 KTEKVDSTQDELSTKCHVGIGVFLDEVYVALHMSCHGSGPFCQSCQHLCTDKYDFTT 1260  
 Db 1214 KTEKVDSTQDELSTKCHVGIGVFLDEVYVALHMSCHGSGPFCQSCQHLCTDKYDFTT 1273  
 QY 1261 HIQGLHRNNAQVEKNGKPK 1281  
 Db 1274 HIQGLHRNNAQVEKNGKPK 1294

RESULT 6  
 AAB92512  
 ID AAB92512 standard; Protein; 785 AA.  
 XX  
 XX AAB92512;  
 XX  
 XX 26-JUN-2001 (first entry)  
 XX  
 XX Human protein sequence SEQ ID NO:10640.  
 DE  
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 KW  
 XX Homo sapiens.  
 OS  
 XX EP1074617-A2.  
 PN  
 XX 07-FEB-2001.  
 PD  
 XX 28-JUL-2000; 2000EP-0116126.  
 PF  
 XX  
 XX 29-JUL-1999; 99JP-0248036.  
 PR  
 XX 27-AUG-1999; 99JP-0300253.  
 PR  
 XX 11-JAN-2000; 2000JP-0118776.  
 PR  
 XX 02-MAY-2000; 2000JP-0183767.  
 PR  
 XX 09-JUN-2000; 2000JP-0241899.  
 XX  
 XX (HELI-) HELIX RES INST.  
 PA  
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto G;  
 PI Iehii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 PI  
 XX WP; 2001-318749/34.  
 DR  
 XX  
 XX Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 XX Claim 8; SEQ ID 10640; 2537pp + CD ROM; English.  
 PS  
 XX  
 XX The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification, where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. the primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX Sequence 785 AA;  
 SQ

Query Match 62.5%; Score 4212; DB 22; Length 785;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 782; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 495 MTKTDKSSSGAKKOPFSSKGAEDNMVTSYNCQFCDFRYSKSHGPDVIVVGPLLRHYQQLH 554  
 Db 1 MTKTDKSSSGAKKOPFSSKGAEDNMVTSYNCQFCDFRYSKSHGPDVIVVGPLLRHYQQLH 63  
 QY 555 NIHKCTIKHCPFCPRGLCSPKHLGEITYPPACRKSNCSCALLLLHLSPGAAGSSRVKH 614  
 Db 6 NIHKCTIKHCPFCPRGLCSPKHLGEITYPPACRKSNCSCALLLLHLSPGAAGSSRVKH 120  
 QY 615 QCHQCSFTTDPVDVLLFHYESVHESQASDVKQEAHNLQSGDQQQSVKSEKHSCTKCFI 674  
 Db 12 QCHQCSFTTDPVDVLLFHYESVHESQASDVKQEAHNLQSGDQQQSVKSEKHSCTKCFI 180  
 QY 675 TOVEEBEISRHRAHSCYKRCQCSFTTAADTQSLLEHFNTHVHCQEQDITTTANGEEDCHALS 734  
 Db 181 TOVEEBEISRHRAHSCYKRCQCSFTTAADTQSLLEHFNTHVHCQEQDITTTANGEEDCHALS 240  
 QY 735 TIKKEPKIDFRVYNLLTPDSKQGEVPSVSVKRELEKCGLEKVKWTESSDDLRNVTM 794  
 Db 241 TIKKEPKIDFRVYNLLTPDSKQGEVPSVSVKRELEKCGLEKVKWTESSDDLRNVTM 800  
 QY 795 RGADILRGSPSYTQASLGLLTPVSGTEQTKLROSQNVVEAAHLARPIYGLAVETKGFQ 854  
 Db 301 RGADILRGSPSYTQASLGLLTPVSGTEQTKLROSQNVVEAAHLARPIYGLAVETKGFQ 360  
 QY 855 GAPAGGEKSGALPQQYPAAGENKSKDESQSLRRRARGSGVFCANCLTTTSLWRKNANGG 914  
 Db 361 GAPAGGEKSGALPQQYPAAGENKSKDESQSLRRRARGSGVFCANCLTTTSLWRKNANGG 420  
 QY 915 YVCNACGLYOKLHSTPRPLNIKQNGEQIRRTKRLNPEALQAEQLNKOORGSNEEQ 974  
 Db 421 YVCNACGLYOKLHSTPRPLNIKQNGEQIRRTKRLNPEALQAEQLNKOORGSNEEQ 480  
 QY 975 VAGSPLEERSSEDLTESHOREIPLPSLSKYEAQGSLSKSHSAQQPVLSQTLDIHKRYQP 1034  
 Db 481 VAGSPLEERSSEDLTESHOREIPLPSLSKYEAQGSLSKSHSAQQPVLSQTLDIHKRYQP 540  
 QY 1035 LHQIKSPQESTGDPGNSSSVSEKSGSSRGSPIEKYMPAKHPNYPGSPKIEKYQYPL 1094  
 Db 541 LHQIKSPQESTGDPGNSSSVSEKSGSSRGSPIEKYMPAKHPNYPGSPKIEKYQYPL 600  
 QY 1095 FGLPFVHNDFOSEADMLRFMSKYKLSVPGNPHYLSHVPGLPNQCQNYVPYPTNLPHFES 1154  
 Db 601 FGLPFVHNDFOSEADMLRFMSKYKLSVPGNPHYLSHVPGLPNQCQNYVPYPTNLPHFES 660  
 QY 1155 AVGSNDNDIPDLDAIKHSRPGPTANGASKEKTKAPPNVKNEGPLNVVVKTEKVDSTQDELS 1214  
 Db 661 AVGSNDNDIPDLDAIKHSRPGPTANGASKEKTKAPPNVKNEGPLNVVVKTEKVDSTQDELS 720  
 QY 1215 TKVHCGIVFLDEVYVALHMSCHGSGPFCQSCQHLCTDKYDFTTHIQGLHRNNAQVE 1274  
 Db 721 TKVHCGIVFLDEVYVALHMSCHGSGPFCQSCQHLCTDKYDFTTHIQGLHRNNAQVE 780  
 QY 1275 KXGK 1278  
 Db 781 KXWK 784

RESULT 7  
 AAB41128

ID AAM41128 standard; Protein; 1213 AA.  
XX AAM41128;  
AC AAM41128;  
XX  
XX 22-OCT-2001 (first entry)  
DT  
XX Human polypeptide SEQ ID NO 6059.  
DE  
DE  
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.  
XX  
XX Homo sapiens.  
OS  
XX WO2001:53312-A1.  
PN  
XX 26-JUL-2001.  
PD  
XX 26-DEC-2000; 2000MO-US34263.  
PF  
XX 21-JAN-2000; 2000US-0488725.  
PR  
XX 25-APR-2000; 2000US-0552117.  
PR  
XX 09-JUL-2000; 2000US-0598042.  
PR  
XX 19-JUL-2000; 2000US-0620312.  
PR  
XX 03-AUG-2000; 2000US-0653450.  
PR  
XX 14-SEP-2000; 2000US-0662191.  
PR  
XX 19-OCT-2000; 2000US-0693036.  
PR  
XX 29-NOV-2000; 2000US-0727344.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX  
XX WPI; 2001-442253/47.  
DR N-PSDB; AAI60284.  
XX  
XX Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
PT  
PT Example 2; SEQ ID NO 6059; 10078pp; English.  
PS  
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM38642-AAAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
XX Sequence 1213 AA;  
SQ  
Query Match 4.2%; Score 286; DB 22; Length 1213;  
Best Local Similarity 19.6%; Pred. No. 2.9e-13;  
Matches 290; Conservative 156; Mismatches 516; Indels 514; Gaps 74;  
QY 48 QSDAAELNHKEBSLH-----VQDP-----SSSSKKDLKS--AVLSE--KAGRYESP 91  
DB 1 QQDFESLADLTDRHRCPCGGDDDDQLSWVASPSSKDVASPTQMGDCD:GLGEEEG 60

QY 92 SKGNGFSPFHDEVTDNKLAFSPAAAGVCEPLKSPQARADQPDMACTPSSGDSLETK 152  
DB 61 QTGLPYPC---QFCXSFIRLSY-----LKRHEQ:HS-DKLPFKCTCYCSRLFKHK 106  
QY 152 EDQKMSPKATEETGAOSGQANCOGLSPVSVASKNPQVPSDGGVRJNKSJTDLLVNDPD 211  
DB 107 RSRDRHIKLA--HTGDKKITHCECAAFS-----RSDHLKHLKTHSSSK 148  
QY 212 PAP-SPELQDFKNC:GGYVYVNDPDTLHKFR-----KYHLGLHNRT----- 254  
DB 149 P-----FKC-VCKRGF--SSTSLSQSHQAQKXG:KEHLAKSEKAKKQDEWCDYCE 197  
QY 255 ---PCDAELDSKILALHNNGVCFSHSKDFCKVN----- 283  
DB 138 DTFSQTELEKHLVLRH--PQISEKADLCQICHPFVDENTLLAH:HQAHANOKHKCPM 255  
QY 284 ---RSVPFSGVLODINSR-----PVL-----LNGTYD---VOVTSQGT----- 315  
DB 256 CPESQFSSVEGVYCHLDSHRQPDSSNHSVSPVPLGVSASKSSATPDSASVERGSTPDST 315  
QY 316 ---F-GIGRKTPOCGNTK-YFRCKFCNFTVMGNSSTELEOHFLOTHPNKTKA----- 364  
DB 316 LKPLRGQKKPRDDGGWTKVYVSCPYCKRDF-NSLAVLE-HLXTIEADKPKQSHTCQIC 374  
QY 365 ---S-PS-----SEVAKPSEKSNKS:PALQSC-----SGDLGKWQCK:TVK-A 435  
DB 375 LQSMPTLYLNNEHVRKLRH--NHAYPVWQFCN:SAFRCNVCFEMFADINSQEHIRVSHC 432  
QY 436 GDDTPVGVSVPIKPLDSRPQNGTEATSVYCKFKGFCSPCESSSS:KJLEHYGKQKQAGVQSG 465  
DB 433 GPNA-----NFSQGNNAFCNQCSMGFLTESS--LTHEICQACHSVGSA 474  
QY 466 GLN-----PELNDKLSRGSVI-----NQNDL-----AKSSEGETMT 496  
DB 475 KLESPVQVQTSFMEVSCPYCTNSPIFGSILKLT:KHKENHKNIPL:AHSKSKABQSPV 534  
QY 497 KTDKSSCAKKKPSKGAEDNMVTS--YNCQFCDFYS-----KSH----- 536  
DB 535 SSOVEVSSPKFORLS---ASANSISNGEYPCNQCDLKFSPNFESFCTHLKJHLELLLRKQA 591  
QY 537 ---GPDVIVVGPLLRHYCQLHNIHKCTIKHCPCFCPRGLCSP---EKHLGEITVYFACRK 589  
DB 592 CFQCKEDFDSQESL:QHL-TVHYMTTTHYVYCESCDKQFSSVDDLQKHLDMPLPLC--- 647  
QY 596 SNGSKCAL-----LLJHLSFGAAGSRVGHQCHQCSFTTPDQDVLLFHYESVHES 639  
DB 648 --CTHCTLCQEVFDSKVSIOVHLAVQHSNEKKN-YRCTACKW----- 686  
QY 646 QASDVKQEAHLOGSDGQGVKES-----KEHSCTKCFITQVEEBEISRHYRAHSCYK 693  
DB 687 ---DFRKEA-----DJCVHYKSHLGNPAKAHKCIFCGETFTSEVELOCHITTHSKYKN 737  
QY 694 CRQCSFTAADTQSJJEHFNTVHCOEQDITTANGEEDG-HAISTIKKEPKIDFRVYNLLTP 752  
DB 733 CKFCSKAPHAILLEKULREKHC-VFDAATENGANGVPPMATKKAEP-ADLOGLMJLKNP 795  
QY 753 DSKMGPEVSVSVKREKLEEKDGLKEKVTWTESSDDL-RNVWTRGADILRGSPSYTQASL 811  
DB 796 EA---FNSH-----EASEDDVDASEPMYGCDDIC--GAAYTMSEVL 829  
QY 812 GLLTPVSGTQFQTKTLDSPNVAAHLARPIYGLAVETKGFLOGAPAGKXSGALPQCYP 871  
DB 810 -----LQNHRLRD-----HNIRP-----GEDDGRKKAEF 854  
QY 872 ASGNKSKQESCSLLRRRRRGSGVFCANCLTTKTSLWPKNANGVYVCNAGLYQKLHSTPR 931  
DB 855 :KSHKNCVCSRTF-----FSENGLAHLOTHRGPAK-HYKCPICG-----EAFPS 899  
QY 932 FLNI:KQNGEQIIRRTKRLNPEALQAEQ:ANKQKQGSNEEQVNGSPLERRSECHLTES 991  
DB 900 LLTUTEH-----KVTHSKS:DTGTCRICKMPLQ--SEEEF:EH 935  
QY 992 HCRFIPLSLSKYAQSLT--KSHSAQQPVLSQT:DIHK--RMQPF:HIQIKSPQESTG 1047

```

Db 936 COMH---PDL-----RNSLTGFRVCVMQVSTSTJELKHGTFRMOKL----- 975
Qy 1048 DPGNSSSVSEKGGSSERGSPKIEKYMRAKHPNYPSPGSPKIEQYQVPLFGLPFVHNDPQSE 1107
Db 976 --AGSSAAS-----SPNGOGLQK-----LYKCALCCKEFSK 1005
Qy 1108 ADWLRFSWKYKLSVPGNPHYL-----SHVPGI--PNPQNYVPYPTFNLPPHFA 1155
Db 1006 QDLV-----KLDVNGLPYGLCAGMARSAHQVGLAPPADR--PCAGLRCEGCVK 1057
Qy 1156 VGSNDIPDLAIKHSRPGPTANGASKEKTKAP--PNVKNQEPNVNVKTEKVDRTQ----- 1210
Db 1058 FESAEDLESHMQVDHRLJTPETSGRGTQTSVPVRKTYQCIKQMTFENEREIQHVA 1117
Qy 1211 -----DELSTKCVHCGIVFLDEVMYALHM-----SCHGDSGPPQCSIC-----QHL 1251
Db 1118 NHMIEGINHECKLQNMQFDSAPKLLCHLIEHSFEGMGGTFCPCVCTFVQANKLOH: 1177
Qy 1252 -----CTDKYDFTTHIQRLHNNQA 1272
Db 1178 FAVHGOEDKIYDCSQCPQKFFFTQELQNTHTMSQHAQ 1213

RESULT 8
ABB64656
ID ABB64656 standard; Protein: 758 AA.
XX
AC ABB64656;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 20760.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
FN WC20017:042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR N-PSDB; ABL08759.
XX
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure; SEQ ID NO 20760; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX

```

```

SQ Sequence 756 AA;
Query Match 4.0%; Score 271; DB 22; Length 758;
Best Local Similarity 19.7%; Pred. No. 2; I.e.12;
Matches 156; Conservative 97; Mismatches 301; Indels 248; Gaps 29;

Qy 551 QQLNHHTKTIKPCPCRGCLCSPEKH-----LGIITPFACRKNCNCSCALJLLHLSPG 605
Db 115 QQHHEHHHLMGGFNPLTPPELNPVQHFFYGGNLRSPQPTPTASATIAPVAV----- 166
Qy 606 AAGSRVKVQCHCCSPTTPDQVLLPHYSVHESASDVKQEAHLQGSDDGQGVK---S 662
Db 167 AGSS-----EKQALTPPMQV--PPASPAKSSQSN:EPEKEHDQMSSEDKYKMAE 218
Qy 663 SKEH-----SCTKCDFTT 675
Db 219 SEDDDTNIKPIYNHSGMKNYKCKTCGVVAITKVDFMAHTRTHMKPKILCCPKCFVT 278
Qy 676 QVEEEISRYR--RAHSCYKQCQCSFTAAQTOSLLEHNT-----VHCQEQIIT-- 723
Db 279 EFKHLEYHIRKHQKQKFCQDKCSYTCVNMJANSHRMSHSSVYQYRCACQDIATKYCH 338
Qy 724 ---ANGEEDGHALSTIKEE-----PKIDFRVTNLT--PDSKMGEPVSESVVKRELEKD 774
Db 339 SFKJLHURKYGHKPGMVLDEGTFNPSLVIDVYTRGPKSKNGGPIASG----- 387
Qy 775 GLKEKVTSSSDDLRNVTWRGADILRGSPSYTQASLGLLTPVSGTQEQTKTIRSPNVE 834
Db 388 -----QSGSGS-----RKSNAVAAPQOQOQSPA--QFVA-TSQLSAAALCGFPLVQ 430
Qy 835 ---AAHLARPYYGLAVETKGFLOGAPAGKEKGA--PCQYPASGENKSKCESOLLARRRG 891
Db 431 GNSAPPAAGSVLPJ-----PASPAKSVASVECTPSLPSP----- 464
Qy 892 SGVFCANCTTKTSWRKANGVYCNACGLYOKLHSTPRNLNIKQNGEGIIRRTRK 951
Db 465 -----ANLPLPLASLLQQRNMAFT-----PWNLN-----LQMLAAQOCAAVAL-----A 504
Qy 952 RLNFEALQAEQLKCKQSGNEFCVNGSPLEKRSSECHLTESHQREIPLPSLSKYEAQCSLT 1011
Db 505 QLSPRMREQLQQNQCCSDNEEEQDDEYERKSVDSAMDJ-----SGTFTV 550
Qy 1012 KSHSAQOPVLVSQTLDIHKRMOPJHIQIKSPQESTGDPQNSSVSEGGKSGSERGPPEKY 1071
Db 551 KEDEQQO-----QPCQLAVNLKVEEAT--PLMSSSNASRRKQ---RVLKLDLT 595
Qy 1072 MRPAKHPNYS-----PGSPLEKYQVPLFG---LPEVNDPQSEADWLEFWSKYKLSVFG 1123
Db 596 LQLRSEATSPQCKVPTPTASPIAGRPXPEEHCSTSSAD----- 641
Qy 1124 NPHVLSHFGLRNFCCQYVVEYPTFNLPPHFSAVGSDNDIFLDLAIKHSRPGPTANGASKE 1183
Db 642 ESMETARVFCANTSASSTA-----SSSGNSSNASSNS 673
Qy 1184 KTKAPNVNKEBPLNVVTKYVDRTQDELSTCKVHCIGIVFLDEVYALHMSCHGDSGPF 1243
Db 674 KNGSSSNSSNGTTSAAVAPPSTPAAGATYECKYCDIFFKDAVLYTHMGVHSCDDVF 733
Qy 1244 QCSICQHLCTDKYDFTTHIQRLH 1267
Db 734 KCMCKEKCQGVGLFVHMARNAH 757

RESULT 9
ABU11707
ID ABU11707 standard; Protein: 499 AA.
XX
XX ABU11707;
XX
XX 13-FEB-2003 (first entry)
XX
XX Human MDET polypeptide SEQ ID 654.
XX

```







```

QY 48 QSDAAELNHHKEHGLH-----VDDP-----SSSKKDLKS--AVLSE--KAGFNYESP 91
Db 13 QODFESLAULTDRAHRCPCGGDDDPQLSWVASPSKDVASPTQMIGDCGLGGBEEG 72
QY 92 SKGDNFSPFHDEVTDRNMLAFSPAAGVCEPLKSPQRAEADDPQDMACTPDSGDSLETK 151
Db 73 GTGLPYPFC---QCCKSFRLSY-----LKRHEQIHS-DKCFPKCTYCSRLFKKH 118
QY 152 EDQKMSPKATEETCCAGSGOANCCGLSPVSVASKNPQVPSDGGVRLNKSXTDLLVNDNDP 211
Db 119 RSRDRHIKL--HTGDKKYHCECAAFS-----RSDHLXILHLKTHSSSK 150
QY 212 PAPLSPBLQPKFCNICGYGYNDPTDLIKHR-----KYHLGLHNRT----- 254
Db 161 P-----FKCTVCKRGF--SSTSLQSHMQAHHKKNKEHLAKSEKAEKKDDFMCDYCE 209
QY 255 ---RODAELOSKILALHNYQFHSKDFOKV-----NSLAVLEIHLKTIHADKPPQGSHTQCIC 356
Db 210 DTPSQTELEKHLVLRH--POLSEKADLQCHCEPEVFVDENTLLAHTHQAHANQKHKCPM 267
QY 284 ---RSVFSGLVDINSR-----PVL-----LNGTYD--VQVTSGGT----- 315
Db 268 CPEQFSSVEGYVCHLDSHRQDSSNHSVSPDVLGVSASMSSATPDSASASVERGSTPDS 327
QY 316 ---FISGRKTPDCGNTK-YFRCKFCNFTYMGNSSTELSOHFLQTHPNKIKKA----- 364
Db 328 LKPLRGQKKMRDGGGWTKVYSPCYSKRDF-NSLAVLEIHLKTIHADKPPQGSHTQCIC 356
QY 365 --SLPS---SEVAKPSEKSNKSI PALQSSD-----SGDLGKWKQKITYK-A 405
Db 387 LDSMPTLYNLNHEHVRKLUK--NHAYPVQFCNLSAFHCNVCPEMFADINSLOEHIRVSHC 444
QY 406 GDDTPVGVSPVKPLDSSRQNGTEATSYMCKFCFSCESSSLKLEHYCKQKHGAVCSG 465
Db 445 GPNA-----NPSDGNNAFFCQCSMGFLTSS--JTEHIQCAHCSVGA 486
QY 466 GLN-----PELNDKLSRGSVI-----NQNDL-----AKSEGETMT 496
Db 487 KLESFVVQPTOSFMVEVSCPYCTNSPIFGSILKLTIKHKNHKNPLAHSKSKAEQSPV 546
QY 497 KTDKSSGAKKKDFSSKGAEDNMVTS--YACOFDFRYSKSHGPDVIVVPLLRHYQCLH 554
Db 547 SSDVEVSPKQRULS---ASANSISNGEYPCNQCDLKFSPFES-----FQTHLKLH 594
QY 555 NIHKCTTIKHCFPCPRGLCSPKEHLGEITYPACRKSX--GSHC-----ALLL 600
Db 595 LELLRLKQACPOCKEDFDSQBSLLQHLTVMYMTTS:HYVCECDKQFSSVDDLLQKHLLOX 654
QY 601 HLSPGAAGSRVHKHQHCQCSFTTPDPDVLVLFHYESVHESQAS-----DYKQEAHL 651
Db 655 H-----TFVLVHCTLCOEVPDSKVSIOVHLAVRHSNKKMYRCTACNWDFRKEA--- 703
QY 652 QGSDGQSVKES-----KEHSCTKCDFITQVEBEISRHYRRAHSCYKRCQCSFTAACTQ 705
Db 704 ---DLQVHVHSHLGNPAKAHKCIFCGEFTSTVELOCHITTHSKVYKRCFKSAFHA 760
QY 706 SLLEHFNTHVHQEQDITITANGEDG-HAISTIKKEPKIDFRVYNLLTPDSKNGEPPVSESV 764
Db 761 LLEKHLREKHC-VFDAATENGANGVPPMATKKAEP-ADLOGLYLLKXPEA----PNSH-- 812
QY 765 VKREKLEKDLKELKVTWESSDDL-RNVYWRGADIIRGSPSYTQASLGLLTPVSGTCEQ 823
Db 813 -----EASEDDVDASEPMYGCIDC--GAAYTMEVJ-----LQ 842
QY 824 TKTLRSDPNVEAHLARPIYGLAVETKGFLOGAGGKSGALPQQVPASGENSKSDSQ 893
Db 843 NHRURD-----HNRIP-----GEDDGRSKKAEPFKGSHKCNVCSR 877
QY 884 SLLRRRSGGVFCANCILTTKSLWRKNANGGVYCNACGLYQKCHSTPRPLNI-KQNGEQ 943
Db 878 TF-----FSENGLEHLOTHRGPAK-HYMCPIG-----ERFSLTLTJTEH----- 917

```

## RESULT 12

```

ABBY1878
ID ABBY1878 standard; Protein; 734 AA.
XX
AC ABBY1878;
XX
DC 26-MAR-2002 (first entry);
XX
DE Drosophila melanogaster polypeptide SEQ ID NC 42426.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
  pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN NC0260171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US03231.
XX
PR 22-MAR-2000; 2000JUS-191537P.
PR 11-JUL-2000; 2000JUS-0614150.
XX
PE (PEXE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
WI KPI; 2001-656960/75.
DR N-PSDB; ABL1598.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
  genes from Drosophila and for elucidating cell signalling and cell-cell
  interactions .
XX
PS Disclosure; SEQ ID NC 42426; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
  capable of detecting 1000 or more genes from Drosophila. The invention is
  useful in developmental biology and in elucidating cell signalling and
  cell-cell interactions in higher eukaryotes for the development of
  insecticides, therapeutics and pharmaceutical drugs. The invention
  discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
  sequences (AB131840-AB1316175) and the encoded proteins

```

```
CC (AB57737-ABB2072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from wipo
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 734 AA;
Query Match 3.7%; Score 256; DB 22; Length 734;
Best Local Similarity 19.4%; Pred. No. 3.3e-11;
Matches 175; Conservative 118; Mismatches 303; Indels 304; Gaps 37;
QY 444 ESSSLKLEHYGKHGAVQSGGLNPELNDKLKLSGVSINQNDLAKSSGEMTKTDKSSS 503
DB 7 KGNHLKASELYGTCH--ERTGG-----KFGQRIYVQWE-----QDTQQQQQQQ 50
QY 504 GAKKDFSKGAEDNMVTSYQCQCD-FRYSKSHGPVIVVGPV-----LRHYQLHNIH 557
DB 51 QQQQQQQQLQQQQHQALTKQQLQLDKIKLESSNGADQLAQOATANNLDEQEQQQHQQQ 110
QY 558 KTIKH--CPTCPGLGSPERHILGITYPFACRK-----SYCSHCA 596
DB 111 AATSVGVVVTQAGAVSEPEEQY--VVVPRNORILLT--AGTLENEAREGEPPS--NAGNA- 167
QY 597 LLLLHLSPGAGSSRVKHQ--CHQCSFTTPDVVLLPHVESVHESQAS-DVQKQANHLQG 653
DB 168 -----SSGSASDSHIEVQSAHQSPGATHVQMAPRNAEVTQGEATGESQQQQQQQQ 221
QY 654 SDGQQ-----SVKSEKHSCTKCDFTQVEEIESRHYRRARHSYKVC 694
DB 222 HQHQQQQQCTPNGASYGETTIVISSEAEALQH-----HHQQQQHQQQCHQCHQHQCA 275
QY 695 RCCSTFADTQSL--EHFNTVHCOEDI--TTANGEDGHAISTIKEPKIDFV----- 746
DB 276 AAASAAAQTVHIAATSHSGTVRFVTEVDFRTAGPETS--ASNMYVDVVPVDSGVHANE 323
QY 747 -----XNLLTFDS-----KMGEPV-----SESVKREKLEEK 773
DB 334 SKTYADLGNAYFPFSSFSFSSNSYAATLQGNITIVYPTGQFLAKSESGJCTGLRQ 393
QY 774 DGLK-----EKWTESSDDLRYNTWR-----GADILGSPSYQAS 810
DB 394 TGPATFQITISFEGNGIEPLMASPAPEYQVCFNPFHPQVIDEYSGNMTSHKWPASS 453
QY 811 LCLLTFVSGTQECTYTLRDSNPVEAHLARPIYGLAVET-----KG----FLOQAFAG 860
DB 454 IQQY-----DGLSVTASTSYSPNHELKNCNCHGPFRLKGSSEYFCNCP--- 496
QY 861 EKSGLPQQYPASGENKSKDESQSLLRRRRSGVFCANCLTKTSLWRKNANGGVVCNAC 920
DB 497 ----AFMRMAPRITQRAKPKAAAPNNR--NGVTCANQINSTLWRNNEGNPVCNAC 551
QY 921 GLYQKLHSTPRPLNTIKQNGEQIITRRTRKRLNPEALQAEQLNQQQRGSNEEQVSGPL 980
DB 552 GLYKLHNMNRLPSKKEG-----IQKRRKPKNN-----GGAPM 596
QY 981 ERSSEDLHTEHQREIPLPSLSKYBAQGSLLTKSHSAQOQVLVSQTLDIHKRQMLHIQIK 1040
DB 587 HR-----APLPSMS-----QGVNLMANSLYPSQVPSVM-----ENSLN 622
QY 1041 SPQES-----GDPCNSSSVSEGGSGSRGSPIEKYMRPAKHPNYSPPGSPIERKY 1091
DB 622 SQONSSELPHDMSITGQAGGQVVISLNAATPPTDGLNMSARHHVTGSHSPYSQOS 681
QY 1092 YPLFLGLPFVHDFQSEADWLRFWSKYKLVSQNPVPHYLGHVPLGNPCQNYVPYPTFNLP 1151
DB 682 TP-----SQOS-----PHLPQ----- 692
QY 1152 HFSAVGSONDIPLDLAIKHSRPGTANGASKETKAPNVNKGELNVVVKTEKVDRSQD 1211
DB 693 -----TVPINRQI--VQVPTIESSRSNTSLTPSV-----ITRTGLPERSNN 734
```

RESULT 13

```
AA093979
ID AA093979 standard; Protein; 1311 AA.
XX
AC AA093979;
XX
DC 13-NOV-2001 (first entry);
XX
DE Human stomach cancer expressed polypeptide SEQ ID NO 27.
XX
KW Human; stomach cancer; marker; screening; micro-metastasis;
peritoneal dissemination.
XX
OS Homo sapiens.
XX
PN WO2000:09317-A1.
XX
PD 08-FEB-2001.
XX
PF 28-JUL-2000; 2000WC-CP05063.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 18-OCT-1999; 99US-0159590.
PR 11-JAN-2000; 2000JP-0118776.
PR 17-FEB-2000; 2000US-0183322.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELIX) HELIX RES INST.
XX
PI Cta T. Isozaki T, Nishikawa T, Hayaashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T, Aburatani H;
PI Kodama T, Midorikawa Y;
XX
XX WP1: 2001-570287/64.
DR N-PSDB; AA193855.
XX
PT New Stomach cancer-associated genes, useful as markers in blood tests
for screening for the early stages of the disease.
XX
PS Claim 1; Page 91-93; 242pp; Japanese.
XX
CC The invention relates to stomach cancer-expressed genes
CC (AA193842-AA193979) and the encoded proteins (AA193967-AA194039). The
CC genes can be used as markers in blood tests for screening for the early
CC stages of the disease. The proteins and peptides can be used as targets
CC for screening for compounds to treat the disease. They can also be used
CC for predicting micro-metastases. The gene can predict peritoneal
CC dissemination.
XX
SQ Sequence 1311 AA;
Query Match 3.7%; Score 256; DB 22; Length 1311;
Best Local Similarity 19.1%; Pred. No. 8.9e-11;
Matches 293; Conservative 178; Mismatches 491; Indels 570; Gaps 80;
QY 19 QLEPFG--TESKVSNNKEFADQMSNTDSDAAELNKEEHLHVQCP-----SSS 70
DB 54 QVFESLSDITEHKI-----NQCLTDGVGVDED-----DPTCSWPASSP 91
QY 71 SKKDLKSAVLSKAGNYESPSKGNFPFPHD-EVTDRLNMLAFSPAAAGGVCEPLKSPQ 129
DB 92 SKDKQTSFSGHSGCDGEE---EGG--POLPFCPCCKSFSLSV-----LKHHE 137
QY 130 RAEDADPQDVACTPSSGDSLETKEDEKMSPKAEETGQAQSGQANCOG-----LSPV 180
DB 138 QSHS-SKLFPKCYCSRLFHKHESDRHKL--HTGDKKYHCSECDAAFRSDHLKILK 194
QY 181 SVASKNP---QVPSDGGVRLNKSITDLLVND-NPQAPLSPQLQDF-----KCNICGYG 230
DB 195 TH-SNKPYPKCAICRRGFLSSSSLHGQMQVHERNKGSGSGRVEDWKMMDTKKCSQCEEG 254
QY 231 YVGNQDPTDLIKHFKYHGLHNPTRQCA-----ELDSKILALHNMVQFSKSKQFQKN 293
```





XX Claim 1a; Page 22-23; 29pp; Japanese.  
XX This invention describes a novel isolated Eos protein and its  
XX encoding nucleic acid. The Eos gene can be used for the development  
XX of new drugs for various cerebral nervous diseases. Independent claims  
XX are also included for the following: (1) a cloned cDNA carrying the  
XX above Eos gene and exhibiting an activity of inhibiting formation of  
XX abnormality in nervous system; (2) an Eos gene expression product having  
XX the above amino acid sequence; (3) a cloned cDNA expressing Eos protein.  
XX its fragment, its derivative and its homolog; (4) a human homolog of Eos  
XX gene; (5) an expression product of the human homolog of Eos gene; (6)  
XX inhibiting formation of abnormality in nervous system including a step of  
XX introducing Eos gene in a cell by using the above cloned cDNA; (7) a drug  
XX composition, that is a gene therapy agent, containing the above Eos gene  
XX as the active component. This represents the mutine embryonic brain Eos  
XX protein described in the method of the invention.  
XX SQ Sequence 533 AA;  
Query Match 3.7%; Score 251.5; DB 21; Length 533;  
Best Local Similarity 18.8%; Pred. No. 4.4e-11;  
Matches 154; Conservative 86; Mismatches 244; Indels 337; Gaps 32;  
QY 484 NDLAKSSEGETM---TKTDKSSGAKKKDFSSKGAEDNMVTSYNCOFCDFRYSKSHGPD- 539  
DB 12 SSLEKDSLGAPGVSVTPNSQHSFSPSLANSIXVEM-----YSDSESRKLLGPDE 64  
QY 540 -----VIVVGPLLRHYQQLNIHKCTIKHCFPCPRGLC---SPEKHLGEITYPFACR 568  
DB 65 RLDDKDSVIVEDSLSE-----PLGYCGSGGPEPH----- 94  
QY 589 KSNCSHCALLLLHLSFGAAGSRVRKHQCQCFTTPDQVLLFHYES-----VHESQ-A 641  
DB 95 -----SPGIRLPNGKLCDCVGMVCGIPNVLVMVHKRSHTGERPFHCNQCQ 640  
QY 642 SDVKQKAN---HLQSDGQGSVKESKEHSTCKDFITQVEEESRHYRAHSC----- 691  
DB 141 ASPTQKGNLLRHKLKASGEKPFK-----CPFCNYACRRRDALTGHL-RTHSVSPTVGK 193  
QY 692 -YKCRQCSFTAAQTQS:LEHFNTVHCQEQDITTANGEEDGHAISTIKKEPKIDFRVYN:LL 750  
DB 194 PYKNCYGRSVKQGSTLEEHEKCHNYLQSLT-----DAQALT----- 232  
QY 751 TPDSKMGEPVSESVVRKELEKDKLKEKVMTESSDDLRNTVTRGADILRGS---PSYT 807  
DB 233 -----GQP-----GDEIRLEWVPDSMLHPSTERPTF- 260  
QY 808 QASLGILLTPVSGTCGQTKTLRDSNVVEAAHLARPIYGLAVETKGFLOGAPAGGEKSGALP 867  
DB 261 DRLANSLT-----KRRRSTP 275  
QY 868 QQYPASGENSKDBESQSLRRRRSSGVFCANCLTTKTLWRKNANGGVYVCNAGLYQKLH 927  
DB 276 QKP--VGEKQMRFSLSL-----PYDVNASGGY----- 301  
QY 928 STPRPLNIKQNGEQIIRRTTRKRLNPEALQAEQLNKQORGSENEOVNGSP:ERRSECH 987  
DB 302 -----EKDVELVAHHGLEPGF-----GGSLAFVGTET 328  
QY 988 LTESHQREIPLPSLUSKYAOGSLTKSHSAQQPVLVSQTLDIHKRMQPLHIQIKSPOESTG 1047  
DB 329 L-----RPLRLPP-----TNCISELTPVISS-----VYTMQPIPSRLSLP----- 364  
QY 1048 DPGNSSSVSGKSGSRGSPTEKYMRAKHPNYSPPGSP:EKYQYPLFG:PFVHNDFOSE 1107  
DB 365 ---GSRAGEGPELDGGGPL---LYRARGSLTDPGASP-----SNGCQDS 404  
QY 1108 ADM:RFWSKYKLSVPGNPHYLHVPGLPNPCQNVYPPTFNLPPIPSAVGSDNDIPLDLA 1167  
DB 405 TD---TESNHEDRIGG-----VGIPSSGSPQFPPTIVVGHSPAYAKEDPKQSG-G 452  
QY 1168 IKHSRPGPTANGASKKXTKAPPNVNKEGPNLVKVKTEKVDVRSIQDELSTKCVHCGIVFLDE 1227

DB 453 LLRTGPGP-----SKEVLRV-----VGSSEFVKAFKCEHCISIFLDH 490  
QY 1228 VVVALRKSHCHSGPFCSCICOHLCCTKYDFTTHIQRLHR 1268  
DB 491 VXFTHKCHGFRDPPFCNICGYSQRYEFSSHIVRGEHK 531  
Search completed: October 29, 2003, 1:138:47  
Job time : 465 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 11:29:57 ; Search time 26 Seconds  
(without alignments)  
2084.623 Million cell updates/sec

Title: US-09-702-216-2

Perfect score: 6851  
Sequence: 1 MVRKNPPLNRVASEGEGQI.....IQRLHNNNAQVKNQKPRE 1281

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PC\*US\_COX8.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	248.5	3.6	516	3	US-08-711-417C-202
2	238	3.5	461	3	US-08-711-417C-196
3	238	3.5	461	4	US-08-733-622C-25
4	233.5	3.4	518	3	US-08-711-417C-196
5	233.5	3.4	518	4	US-08-733-622C-27
6	230.5	3.4	568	5	PCT-US93-08743-5
7	225	3.3	334	3	US-08-711-417C-201
8	225	3.3	334	4	US-08-733-622C-23
9	224.5	3.3	432	3	US-08-711-417C-197
10	224.5	3.3	432	4	US-08-733-622C-26
11	210.5	3.1	376	3	US-08-711-417C-200
12	210.5	3.1	376	4	US-08-733-622C-29
13	209.5	3.1	232	4	US-08-733-622C-31
14	208	3.0	236	3	US-08-711-417C-175
15	206.5	3.0	390	3	US-08-711-417C-199
16	206.5	3.0	390	4	US-08-733-622C-28
17	205	3.0	431	3	US-08-711-417C-195
18	205	3.0	431	4	US-08-733-622C-24
19	200	2.9	507	4	US-08-733-622C-2
20	200	2.9	976	3	US-08-894-997-50
21	195.5	2.9	236	3	US-08-711-417C-174
22	193.5	2.8	853	3	US-09-254-325-2
23	190.5	2.8	238	3	US-08-711-417C-177
24	189.5	2.8	470	3	US-08-465-590-153
25	189.5	2.8	470	2	US-08-283-300A-13
26	188.5	2.8	470	4	US-08-733-622C-22
27	188.5	2.8	470	5	PCT-US95-09345-13

Query Match 3.6% Score 248.5; DB 3; Length 516;  
Best Local Similarity 25.7%; Pred. No. 5.3e-13;

US-08-711-417C-202  
; Sequence 202, Application US/08711417C  
; Patent No. 6228611  
; GENERAL INFORMATION:  
; APPLICANT: Georgopoulos, Karia A.  
; TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE  
; NUMBER OF SEQUENCES: 202  
; CORRESPONDENCE ADDRESS:  
; ADDRESSES: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: PastSeq for Windows Version 2.0b  
; CURRENT APPLICATION NUMBER: US/08/711-417C  
; FILING DATE: 05-Sep-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/238,212  
; FILING DATE: 02-MAY-1994  
; APPLICATION NUMBER: 09/121,438  
; FILING DATE: 14-SEP-1993  
; APPLICATION NUMBER: 07/946,233  
; FILING DATE: 14-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Myers, Louis P.  
; REGISTRATION NUMBER: 35,965  
; REFERENCE/DOCKET NUMBER: 10287/007001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 202:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 516 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
; SEQUENCE DESCRIPTION: SEQ ID NO: 202:  
US-08-711-417C-202

ALIGNMENTS

RESULT 1

US-08-711-417C-202  
; Sequence 202, Application US/08711417C  
; Patent No. 6228611  
; GENERAL INFORMATION:  
; APPLICANT: Georgopoulos, Karia A.  
; TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE  
; NUMBER OF SEQUENCES: 202  
; CORRESPONDENCE ADDRESS:  
; ADDRESSES: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: PastSeq for Windows Version 2.0b  
; CURRENT APPLICATION NUMBER: US/08/711-417C  
; FILING DATE: 05-Sep-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/238,212  
; FILING DATE: 02-MAY-1994  
; APPLICATION NUMBER: 09/121,438  
; FILING DATE: 14-SEP-1993  
; APPLICATION NUMBER: 07/946,233  
; FILING DATE: 14-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Myers, Louis P.  
; REGISTRATION NUMBER: 35,965  
; REFERENCE/DOCKET NUMBER: 10287/007001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 202:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 516 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
; SEQUENCE DESCRIPTION: SEQ ID NO: 202:  
US-08-711-417C-202

Query Match 3.6% Score 248.5; DB 3; Length 516;  
Best Local Similarity 25.7%; Pred. No. 5.3e-13;







```

; TITLE OF INVENTION: AIOLOS GENE
; FILE REFERENCE: 10287-030001
; CURRENT APPLICATION NUMBER: US/08/733,622C
; CURRENT FILING DATE: 1996-10-17
; PRIOR APPLICATION NUMBER: US 62/017,646
; PRIOR FILING DATE: 1996-05-14
; PRIOR APPLICATION NUMBER: US 62/005,529
; PRIOR FILING DATE: 1995-10-18
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Mus musculus
; US-08-733-622C-27

Query Match      3.48; Score 233.5; DB 4; Length 518;
Best Local Similarity 23.18; Pred. No. 1.1e-11;
Matches 126; Conservative 70; Mismatches 192; Indels 157; Gaps 25;

QY 814 LTPVSGTQEQTKLRD-SPNVEAAHLARPIYGLAVETKGFQCAPA---GGEKSGALPQQ 869
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 37 LSTTSGAQONSKSDRGMASNVK-----VETQSDENGRCACENGECAEDLRM 84
QY 870 YPASGENKS---KDESQSLRRRRRGSGVFCANCLTTKTSLRKNANGGYVCNACGL--- 922
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 85 LDASGERKNGSHRDQGSAL-----SGVGT-----RPNKGKUCDTCGIVC-G 128
QY 923 -----YOKLHSTPRP-----LNIIKQNGEQ-----IIRRTKRLNRP 955
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 129 PNVLWVHKRSHTGTPQCQNSGASFTQKGNLRLHKLHSGEPPFKCHLCNACRRR--- 185
QY 956 EALQAEQLNQKQSGNEQVNGSPLERRS--EDHLTSHQ--REIPLPSL-----SK 1003
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 186 DALTGHRLTHSVGKPHKCGYKORSLSEHKEKCHNYLESMLGPGVCPVKEETNH 245
QY 1004 YEAQGSITKSHSAQPVLSQTLDIHKRMQPLHIQIKSPOESTGDPNSSSVSEKGSSE 1063
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 246 NEMAEADLCKGAERSLVDRLASNAVKRKSSM-----PKFGLDKCLSDMPYDSANYEK 293
QY 1064 RGSPIEYMRPAKHP--NY-----SPPGS-----PIEKYQY-----PLFGPFPVH 1101
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 300 EDMMTSHVMDQAINNAIYNLGAESLRPLVQTPGSSSEVPVISMVQLHKFPDGGPRSN 359
QY 1102 NDFQSEADNLRFWSKYKLSVPCNPHYLSHVPGLP-NFCQNYVP-----YPTF 1147
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 360 HSAQDAVDNLLLSKAK-SVS-----SREASPSNSCODSTDTESNAEQRSGLIYCTN 412
QY 1148 NLPHFSAVGSDNDIPLDLAIKHSRPGPTANGASKEKTKAPPNVKNEGPLNVVKTKEVDR 1207
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 413 HINPH-----ARNGLAKSEQRAYEVLRAASENSQDAFRVYS 449
QY 1208 STQDELST-KCVHCGIVPLDEWYALHMSCHGDSG---PFQCSICQHLCTDKYDFTHIQ 1263
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 450 TSGEOLKYKCEHCVRLFDDHVMYTIHMGCHGCHGFPDPFECNMGVHVSQDRVEFSHIT 509
QY 1264 RGLHR 1268
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 510 RGEHR 514
;

RESULT 6
PCT-US93-08743-5
; Sequence 5, Application PC/TUS9308743
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE
; NUMBER OF SEQUENCES: 152
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08743
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 946,233
; FILING DATE: 14-SEP-1992
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7450
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 568 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; PCT-US93-08743-5

Query Match      3.48; Score 230.5; DB 5; Length 568;
Best Local Similarity 22.18; Pred. No. 2.5e-11;
Matches 128; Conservative 73; Mismatches 23; Indels 175; Gaps 26;

QY 814 LTPVSGTQEQTKLRD-SPNVEAAHLARPIYGLAVETKGFQCAPA---GGEKSGALPQQ 869
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 37 LSTTSGAQONSKSDRGMASNVK-----VETQSDENGRCACENGECAEDLRM 84
QY 970 YPASGENKS---KDESQSL-----IIRRTKRLNPEALQALNQKQSGNEEQVNGSPLE 894
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 85 LDASGERKNGSHRDQGSALSGVGIPLNPKLKCICGVCIGPNVLMVHKRSHTGERP 144
QY 895 F-CANCLTTKTSLR-KNANGGYVCNACGL-----YOKLHSTPRP----- 932
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 145 FQCQCASALSGVGIPLNPKLKCICGIVCIGPNVLMVHKRSHTGERPFCQCCGASF 204
QY 933 -----LNIIKQNGEQ-----IIRRTKRLNPEALQALNQKQSGNEEQVNGSPLE 981
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 205 TQKGNLRLHKLHSGEPPFKCHLCNACRRR---DALTGHRLTHSVGKPHKCGYKRSYK 261
QY 982 PRRS--EDHLTSHQ--REIPLPSL-----SKYEAQGSITKSHSAQPVLSQTLDIH 1029
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 262 CRSSLEEKERKCHNYLESMLGPGVPIKBTNHNMAEDLCKIGAERSLVDRLASNAVA 321
QY 1030 KRMQPLHIQIKSPOESTGDPNSSSVSEKGSSEKSPLEKYMERAHP--NY----- 1080
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 322 KRKSSM-----PKFGLDKCLSDMPYDSANYEKEDMTSHVMDQAINNAIYNLGAESLR 375
QY 1081 -----SPPGS-----PIEKYQY-----PLFGPFPVHNDQFSEADNLRFWSKYKLSVEGNPHY 1127
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 376 PLVQTPGSSSEVVEISSNYGLHKFPDGGPPRSNESAQDAVDNLLLSKAK-SVS----- 429
QY 1128 LSHVPGLP-NFCQNYVP-----YPTNLPHFSAVGSDNDIPLDLAIKHSRP 1173
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 430 -SREASPSNSCODSTDTESNAEQRSGLIYTNHNEH----- 467
QY 1174 GPTANGASKEKTKAPPNVKNEGPLNVVKTKEVDASTODELST-KCVHCG-VFLDEVNVAL 1232
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 468 --ARNGLAKSEQRAYEVLRAASENSQDAFRVYSTGEOLKYKCEHCVRLFDDHVMYTI 525
QY 1233 RYSCHGDSG---PFQCSICQHLCTDKYDFTHIQRLHR 1268
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 526 RMGCHGCHGFPDPFECNMGVHVSQDRVEFSHITRGEHR 564
;

RESULT 7
US-08-711-417C-201
; Sequence 201, Application US/08711417C
; Patent No. 6228611
; GENERAL INFORMATION:
; APPLICANT: Georgopoulos, Katia A.
; TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE
; NUMBER OF SEQUENCES: 202
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.

```

STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/711,417C  
FILING DATE: 05-Sep-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/238,212  
FILING DATE: 02-MAY-1994  
APPLICATION NUMBER: 08/121,438  
FILING DATE: 14-SEP-1993  
APPLICATION NUMBER: 07/946,233  
FILING DATE: 14-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Myers, Louis P.  
REGISTRATION NUMBER: 35,965  
REFERENCE/DOCKET NUMBER: 10287/007061  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 201:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 334 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
SEQUENCE DESCRIPTION: SEQ ID NO: 201:  
US-08-711-417C-201

Query Match 3.3%; Score 225; DB 3; Length 334;  
Best Local Similarity 24.2%; Pred. No. 3e-11;  
Matches 94; Conservative 53; Mismatches 145; Indels 96; Gaps 17;

QY 915 YVCNACGL-----YQKLHSTPRPLNIKON-----NGEQIARRTRK 951  
DB 5 FQNCQGASFTQGNLLRHILKHSGEKPKCHLCNYACRRDALTGHLRTHSVIKEETK- 63

QY 952 RLNPEALQAEQLNK--QQRGSNEEQVNGSPLEERS---EDLHTEHQSEIPLPLSLSKYEA 1006  
DB 64 ----HSEMAEDLCIKGSESLVLDRLASNVAKRKSMFQKFLGDKGLSDTPYDS-ATYEK 118

QY 1007 QGSLTKSHSAQOPVLVSQTLDIHKRMQPLHIQIKSPQESTGPGNSSSVSEKGSSEKGS 1066  
DB 119 ENEMKSHVMQDAINNAINYLGAESLRPL---VQTP-----PG-----GS 155

QY 1067 PIEKYMRPAKHPNYSPPGSPPIEKYQYPLFGLPFVHNDQFSEADMLRFMSKYKLSVPGNPH 1126  
DB 156 EVVPVISMYQLHRRSEGTGRSN-----HSAQDSAVEYLLLSKAKL-VP-----199

QY 1127 YLSHVVGLP-NPCQNVVPYPTFNLPHPFSVAGSNDIP-----LDLA:KHSRPGPTANGA 1180  
DB 200 --SEREASPSNCSQDSTESNNEORSGLIYLTNHIARRAQRVSLEKEH-RAYDYLRAA 256

QY 1181 SKEKTAPPNVKNQEGPLNVVKTTEKVDRTQDELSTKCVHCGIVFLDEVYALHMSCHGDS 1240  
DB 257 SENSQDA-----LRVST-----SGEQMKVYKCEHCRVFLDHWYTIHYGCGEGR 302

QY :241 GPFQCSICQHLCTDKYDFTTHIQRGJHR 1268  
DB 303 DPPECNMGYHSQCRVEYFSSHITRGEHR 330

RESULT 8  
US-08-733-622C-23

Sequence 23, Application US/08733622C  
Patent No. 6528634  
GENERAL INFORMATION:  
APPLICANT: Georgopoulos, Katia  
APPLICANT: Morgan, Bruce A.  
TITLE OF INVENTION: AIOLOS GENE  
FILE REFERENCE: 10287-030001  
CURRENT APPLICATION NUMBER: US/08/733,622C  
CURRENT FILING DATE: 1996-10-17  
PRIOR APPLICATION NUMBER: US 60/017,646  
PRIOR FILING DATE: 1996-05-14  
PRIOR APPLICATION NUMBER: US 60/005,529  
PRIOR FILING DATE: 1995-10-18  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 23  
LENGTH: 334  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-08-733-622C-23

Query Match 3.3%; Score 225; DB 4; Length 334;  
Best Local Similarity 24.2%; Pred. No. 3e-11;  
Matches 94; Conservative 53; Mismatches 145; Indels 96; Gaps 17;

QY 915 YVCNACGL-----YQKLHSTPRPLNIKON-----NGEQIARRTRK 951  
DB 5 FQNCQGASFTQGNLLRHILKHSGEKPKCHLCNYACRRDALTGHLRTHSVIKEETK- 63

QY 952 RLNPEALQAEQLNK--QQRGSNEEQVNGSPLEERS---EDLHTEHQSEIPLPLSLSKYEA 1006  
DB 64 ----HSEMAEDLCIKGSESLVLDRLASNVAKRKSMFQKFLGDKGLSDTPYDS-ATYEK 118

QY 1007 QGSLTKSHSAQOPVLVSQTLDIHKRMQPLHIQIKSPQESTGPGNSSSVSEKGSSEKGS 1066  
DB 119 ENEMKSHVMQDAINNAINYLGAESLRPL---VQTP-----PG-----GS 155

QY 1067 PIEKYMRPAKHPNYSPPGSPPIEKYQYPLFGLPFVHNDQFSEADMLRFMSKYKLSVPGNPH 1126  
DB 156 EVVPVISMYQLHRRSEGTGRSN-----HSAQDSAVEYLLLSKAKL-VP-----199

QY 1127 YLSHVVGLP-NPCQNVVPYPTFNLPHPFSVAGSNDIP-----LDLA:KHSRPGPTANGA 1180  
DB 200 --SEREASPSNCSQDSTESNNEORSGLIYLTNHIARRAQRVSLEKEH-RAYDYLRAA 256

QY 1181 SKEKTAPPNVKNQEGPLNVVKTTEKVDRTQDELSTKCVHCGIVFLDEVYALHMSCHGDS 1240  
DB 257 SENSQDA-----LRVST-----SGEQMKVYKCEHCRVFLDHWYTIHYGCGEGR 302

QY 1241 GPFQCSICQHLCTDKYDFTTHIQRGJHR 1268  
DB 303 DPPECNMGYHSQCRVEYFSSHITRGEHR 330

RESULT 9  
US-08-711-417C-197  
Sequence 197, Application US/08711417C  
Patent No. 6228611  
GENERAL INFORMATION:  
APPLICANT: Georgopoulos, Katia A.  
TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE  
NUMBER OF SEQUENCES: 202  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95



```

STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/711,417C
  FILING DATE: 05-SEP-1996
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 08/238,212
  FILING DATE: 02-MAY-1994
  APPLICATION NUMBER: 08/121,438
  FILING DATE: 14-SEP-1993
  APPLICATION NUMBER: 07/946,233
  FILING DATE: 14-SEP-1992
ATTORNEY/AGENT INFORMATION:
  NAME: Myers, Louis P.
  REGISTRATION NUMBER: 35,965
  REFERENCE/DOCKET NUMBER: 10287/007001
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 617/542-5070
  TELEFAX: 617/542-8906
  TELEX: 200154
INFORMATION FOR SEQ ID NO: 200:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 376 amino acids
    TYPE: amino acid
    TOPOLOGY: linear
  MOLECULE TYPE: protein
  FRAGMENT TYPE: internal
  SEQUENCE DESCRIPTION: SEQ ID NO: 200:
US-08-711-417C-200

```

```

Query Match      3.1%; Score 210.5; DB 3; Length 376;
Best Local Similarity 22.3%; Pred. No. 7e-10;
Matches 107; Conservative 60; Mismatches 144; Indels 169; Gaps 21;

QY 814 LTPVSGTQEQTKLRD-SPNVEAAHLARPIYGLAVETKGFLOGAPA---GGEKSGALPQQ 869
DB 37 LSTTSGAQQNSKSDRGYASNVK-----VETQSDENGRCACENMGCECEDLRLM 84
QY 870 YPASGENKS---KDESQSILRRRGSGVFCANCLT*KTSLMRKNANGGVYVCHACGLYQKL 926
DB 85 LDASGEKMGNSHRDQSSAL-----SGVGGI-----RLPNGKLKCDICGV--- 125
QY 927 HSTPRPLNIKKONGEQIPRRTRKRLNPEALQAEQLNKQQRGSNEEQVNGSPLEERSSED 986
DB 126 -----CIGNVLM-----VHKRS--- 138
QY 987 HLTESHQREIPLPSLSKYEAQGSLLTKSHSAQQPVLVSQTLDIHKRMQPLHIQIKSPQEST 1046
DB 139 HTGDKCLSDMPYDS-ANYEKEDMMT-SHWMDQAINNAIYVLAESLRPL---VQTE--- 189
QY 1047 GDPNSSSVSEGGSGSERGSPIEKYMRPAKHPNYSPGSPFIEKYQYPLFGLFVHNDPQS 1106
DB 190 -----PGSSEVVPVSSMYQLHKPPSDGPPRS-----NHSAGQ 222
QY 1107 EADWLRFWSKYKLSVPCGNPHYLSHVPLGP-NPCQNYVP-----YPTFNLPHP 1152
DB 223 AVNLLLLSKAK-SVS-----SERASPSNSQDSTDTESNAEQRSGLIYVTHINPH 275
QY 1153 FSAVGSNDIPDLAIKHSPGPTANGASKERTKAPPNVKNEGFLNVVTKVCRSTQDE 1212
DB 276 -----ARNGLALKEQRAVEVLRAASENSQDAFRVVSSTGEQ 312
QY 1213 LST-KCVHCCGVFDEWYVALHMSCHGDSG---PFQCSICQHLCTCKYDFTTHIQSLGHR 1268
DB 313 LKVVKCHCRVFLDHVMYTHMGCHGCHGFRDPFCNMGYHSQDRYEFSSHITRGEHR 372

```

```

RESULT 13
US-08-733-622C-31
; Sequence 31, Application US/08/733622C
; Patent No. 6528634
; GENERAL INFORMATION:
; APPLICANT: Georgopoulos, Katia
; APPLICANT: Morgan, Bruce A.
; TITLE OF INVENTION: AIOLOS GENE
; FILE REFERENCE: 10287-030001
; CURRENT APPLICATION NUMBER: US/08/733,622C
; PRIOR FILING DATE: 1996-10-17
; PRIOR APPLICATION NUMBER: US 60/017,646
; PRIOR FILING DATE: 1996-05-14

```

```

RESULT 12
US-08-733-622C-29
; Sequence 29, Application US/08/733622C
; Patent No. 6528634
; GENERAL INFORMATION:
; APPLICANT: Georgopoulos, Katia
; APPLICANT: Morgan, Bruce A.
; TITLE OF INVENTION: AIOLOS GENE
; FILE REFERENCE: 10287-030001
; CURRENT APPLICATION NUMBER: US/08/733,622C
; CURRENT FILING DATE: 1996-10-17
; PRIOR APPLICATION NUMBER: US 60/017,646
; PRIOR FILING DATE: 1996-05-14
; PRIOR APPLICATION NUMBER: US 60/005,529
; PRIOR FILING DATE: 1995-10-18
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Xus musculus
US-08-733-622C-29

```

```

Query Match      3.1%; Score 210.5; DB 4; Length 376;
Best Local Similarity 22.3%; Pred. No. 7e-10;
Matches 107; Conservative 60; Mismatches 144; Indels 169; Gaps 21;

QY 814 LTPVSGTQEQTKLRD-SPNVEAAHLARPIYGLAVETKGFLOGAPA---GGEKSGALPQQ 869
DB 37 LSTTSGAQQNSKSDRGYASNVK-----VETQSDENGRCACENMGCECEDLRLM 84
QY 870 YPASGENKS---KDESQSILRRRGSGVFCANCLT*KTSLMRKNANGGVYVCHACGLYQKL 926
DB 85 LDASGEKMGNSHRDQSSAL-----SGVGGI-----RLPNGKLKCDICGV--- 125
QY 927 HSTPRPLNIKKONGEQIPRRTRKRLNPEALQAEQLNKQQRGSNEEQVNGSPLEERSSED 986
DB 126 -----CIGNVLM-----VHKRS--- 138
QY 987 HLTESHQREIPLPSLSKYEAQGSLLTKSHSAQQPVLVSQTLDIHKRMQPLHIQIKSPQEST 1046
DB 139 HTGDKCLSDMPYDS-ANYEKEDMMT-SHWMDQAINNAIYVLAESLRPL---VQTE--- 189
QY 1047 GDPNSSSVSEGGSGSERGSPIEKYMRPAKHPNYSPGSPFIEKYQYPLFGLFVHNDPQS 1106
DB 190 -----PGSSEVVPVSSMYQLHKPPSDGPPRS-----NHSAGQ 222
QY 1107 EADWLRFWSKYKLSVPCGNPHYLSHVPLGP-NPCQNYVP-----YPTFNLPHP 1152
DB 223 AVNLLLLSKAK-SVS-----SERASPSNSQDSTDTESNAEQRSGLIYVTHINPH 275
QY 1153 FSAVGSNDIPDLAIKHSPGPTANGASKERTKAPPNVKNEGFLNVVTKVCRSTQDE 1212
DB 276 -----ARNGLALKEQRAVEVLRAASENSQDAFRVVSSTGEQ 312
QY 1213 LST-KCVHCCGVFDEWYVALHMSCHGDSG---PFQCSICQHLCTCKYDFTTHIQSLGHR 1268
DB 313 LKVVKCHCRVFLDHVMYTHMGCHGCHGFRDPFCNMGYHSQDRYEFSSHITRGEHR 372

```

; PRIOR APPLICATION NUMBER: US 60/005,529  
; PRIOR FILING DATE: 1995-10-18  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 31  
; LENGTH: 232  
; TYPE: PRT  
; ORGANISM: Gallus gallus  
US-08-733-622C-31

Query Match 3.1%; Score 209.5; DB 4; Length 232;  
Best Local Similarity 24.8%; Pred. No. 3.7e-10;  
Matches 71; Conservative 43; Mismatches 99; Indels 73; Gaps 9;  
QY 995 EIPPLSLKSYAQSGLTKSHSAQPVLSQTLDIHKMOPHIOIKSPQESTGPGKSSS 1054  
DB 4 DLPYDATTNYKENEIMQTHVIDQAINNAISYLAESLRPL---VQTP----- 48  
QY 1055 VSEKGSSESPIEKYMRAKHPNYPSPGSPIEKYQVPLFGLPVFNDFQ----- 1105  
DB 49 -----PVGSEVVPV-SPMYQLHKPHGDNQTRSNHTAQD 81  
QY 1106 SEADWLRFWSKYKLSVPCNPHYLSHVPLGP-LNPNQNYVPYPTFNLPHFSAVGSNDIPL 1164  
DB 82 SAVENLLLSKAK-SVS-----SERDASPSNSCOD-----STDTSNNEERS 122  
QY 1165 DLAIKHSRPGPTA-NGASKEKTKAPPNVKNEGPLNVVKTCKVDSTQDELST-KCVHCGI 1222  
DB 123 GLIVLTNIGHPHANGISVKEESRQFVLVLRAGTQNSDAFKVISSNGEYVYKCEHCV 182  
QY 1223 VFLDEWYALHMSCHDGGPQCSCIQHLCTDKYDFTTHIQRG-LHR 1268  
DB 193 LFLDHVMYTIHMGCHGFRDPFECNMGYHSQDRYEFSSHTIRGEHR 228

## RESULT 14

US-08-711-417C-175  
; Sequence 175, Application US/08711417C  
; Patent No. 6228611  
; GENERAL INFORMATION:  
; APPLICANT: Georgopoulos, Katia A.  
; TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE  
; NUMBER OF SEQUENCES: 202  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: FastSeq for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/711,417C  
; FILING DATE: 05-Sep-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/238,212  
; FILING DATE: 02-MAY-1994  
; APPLICATION NUMBER: 08/121,438  
; FILING DATE: 14-SEP-1993  
; APPLICATION NUMBER: 07/946,233  
; FILING DATE: 14-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Myers, Louis P.  
; REGISTRATION NUMBER: 35,965  
; REFERENCE/DOCKET NUMBER: 10287/007001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154

; INFORMATION FOR SEQ ID NO: 175:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 236 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: Protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 175:  
US-08-711-417C-175

Query Match 3.0%; Score 208; DB 3; Length 236;  
Best Local Similarity 26.6%; Pred. No. 5.1e-10;  
Matches 72; Conservative 35; Mismatches 106; Indels 58; Gaps 9;  
QY 1004 YEAGSGLTKSHSAQPVLSQTLDIHKMOPHIOIKSPQESTGPGSSSVSEKGSSE 1063  
DB 15 YEKENEMKSHVMDQAINNAISYLAESLRPL---VQTP-----PGSEVV----- 57  
QY 1064 RSPIEKYMRAKHPNYPSPGSPIEKYQVPLFGLPVFNDFQSEA-DWLRFWSKYKLSVP 1122  
DB 58 -----PVISPMYQLHRRSEGTFRSNHSAQNSAVEYLLLSKAKLEKK 99  
QY 1123 GNPHYLSHVPLGP-LNPNQNYVPYPTFNLPHFSAVGSNDIPL-----LDLAIKHSRPGPTA 1177  
DB 100 KVPSERESAPS--NSCQSTDTESNNEQRSGLIYLTNHIARRAQRVSLKEEH-RAYDLL 156  
QY 1178 NGASKEKTKAPPNVKNEGPLNVVKTCKVDSTQDELSTKCVHCGIPLFDEVMYALHXSCH 1237  
DB 157 RAASENSQDA-----LWVST-----SGEQMKVYKCEHCVLFLDHVMYTIHMGCH 202  
QY 1238 GDSGFPQCS-COHLCTDKYDFTTHIQRG-LHR 1268  
DB 203 GPRDFECNMGYHSQDRYEFSSHTIRGEHR 233

## RESULT 15

US-08-711-417C-199  
; Sequence 199, Application US/08711417C  
; Patent No. 6228611  
; GENERAL INFORMATION:  
; APPLICANT: Georgopoulos, Katia A.  
; TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE  
; NUMBER OF SEQUENCES: 202  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: FastSeq for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/711,417C  
; FILING DATE: 05-Sep-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/238,212  
; FILING DATE: 02-MAY-1994  
; APPLICATION NUMBER: 08/121,438  
; FILING DATE: 14-SEP-1993  
; APPLICATION NUMBER: 07/946,233  
; FILING DATE: 14-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Myers, Louis P.  
; REGISTRATION NUMBER: 35,965  
; REFERENCE/DOCKET NUMBER: 10287/007001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 199:

```
;
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 390 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   FRAGMENT TYPE: internal
;   SEQUENCE DESCRIPTION: SEQ ID NO: 199:
US-08-711-417C-199

Query Match      3.0%; Score 206.5; DB 3; Length 390;
Best Local Similarity 24.0%; Pred No 1.7e-C9;
Matches 106; Conservative 58; Mismatches 159; Indels 119; Gaps 20;

QY 864 GALPOQYPASGENKKSQESQSLRRRRSGGVFCAN---CLTKTSLWR--KNANGG---914
Db 27 GDEPMVPEDLSLTSGAQONSKSDRMGMRPFQCNQSGASFCKGNLLRHKLHSGEKPF 86
QY 915 --YVCN-ACGLYOKLHSTPRPLNIKQ--NNGEQIIRRTKRRLNPEALQAEOLNK--QQ 567
Db 87 KCHLCNYACRRRDALTGHLRTHSVIKETNHE-----XAEELCKIGAE 130
QY 968 RGSNEEQVNGSPLEERS---EDHLTESHOREIPLPSLKYEAQGSUTKSHSAQAPVLVSC 1024
Db 131 RSLVLRSLASNAVAKRSSMPQKFLGDKCLSDMPYDS-ANYEKEDNMT-SHYWDQALNNAI 188
QY 1025 TLDIHKMOPLHIQIKSPQESTGDPNCGSSSVSEKSGSERGSPIEKYMRPAKHNYSPPG 1094
Db 189 NYLGAESLRPL---VQTP-----FGSSEVVPVISMYQLHKPESDGP 229
QY 1085 SPIEKYQYPLFGLPFVHNDFOSEADWLREWSKYKLSVPCNPHYLGHVPLP-NPCQNVVP 1142
Db 230 S-----NHSAQDAVDNLLLSKAK-SVS-----SEREASPSNSCCDS 267
QY 1144 -----YPTFNLPHPFSAVGSDND:PLDLAIKHSEPGPTANGASKENTKAPPN 1190
Db 268 TESNAEEQRSGLIYLTNH:NPH-----ARNGLAKKEQRAYE 304
QY 1:91 VKNEGPLNVVYKTEKVDNSTCDELST-KCYHCGIVFLDEVNVALHMSCHCDSQ---PFQCS 1246
Db 305 VLRAAENSQDAFRVYSTGEQUKYKCEHCRVLFLDHVMYTIHMSCHGCHGFRDFFECN 364
QY 1247 ICOHLCTDKYDFTTHIORGLHR 1268
Db 365 MCGYHSCDRYEFSSHITRGEHR 386
```

Search completed: October 29, 2003, 11:39:21  
Job time : 28 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 11:39:25 ; Search time 230 Seconds

(without alignments)

953.209 Million cell updates/sec

Title: US-09-702-216-2

Perfect score: 6851

Sequence: 1 MVRKNPPLRNVAESGGQT.....IQRLHRRNAQVEKNGKPKR 1281

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 642050 seqs, 171146064 residues

Total number of hits satisfying chosen parameters: 642050

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

```
1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6851	100.0	1281	15	US-10-177-293-472 Sequence 472, App
2	246	3.6	537	14	US-10-037-667-1 Sequence 1, Appl
3	242	3.5	515	14	US-10-037-667-4 Sequence 4, Appl
4	240	3.5	419	15	US-10-037-823-321 Sequence 321, App
5	238	3.5	461	10	US-09-755-830-38 Sequence 38, Appl
6	233.5	3.4	518	10	US-09-755-830-40 Sequence 40, App
7	225	3.3	334	10	US-09-755-830-43 Sequence 43, Appl
8	224.5	3.3	432	10	US-09-755-830-39 Sequence 39, App
9	222	3.2	323	10	US-09-764-864-947 Sequence 947, App
10	222	3.2	323	10	US-09-764-864-1095 Sequence 1095, Ap
11	217.5	3.2	294	10	US-09-764-864-1515 Sequence 1515, Ap
12	215	3.1	442	10	US-09-749-728B-11 Sequence 11, Appl
13	211.5	3.1	1125	10	US-09-974-298-114 Sequence 14, App
14	210.5	3.1	376	10	US-09-755-830-42 Sequence 42, Appl
15	209	3.1	474	15	US-10-177-293-174 Sequence 174, App

16	209	3.1	2783	10	US-09-816-669A-14 Sequence 14, Appl
17	207.5	3.0	757	15	US-10-014-789A-4 Sequence 4, Appl
18	207	3.0	443	15	US-10-177-293-176 Sequence 176, App
19	206.5	3.0	390	10	US-09-755-830-41 Sequence 41, Appl
20	206	3.0	498	14	US-10-037-667-5 Sequence 5, Appl
21	205	3.0	431	10	US-09-755-830-37 Sequence 37, Appl
22	204	3.0	517	12	US-10-032-585-7874 Sequence 7874, Ap
23	201	2.9	507	14	US-10-037-667-3 Sequence 3, Appl
24	198	2.9	532	14	US-10-037-667-2 Sequence 2, Appl
25	196	2.9	481	10	US-09-764-864-973 Sequence 979, App
26	196	2.9	1254	15	US-10-232-561-10 Sequence 10, Appl
27	195	2.8	876	10	US-09-801-368-44 Sequence 44, Appl
28	194.5	2.8	135	10	US-09-764-864-983 Sequence 983, App
29	193.5	2.8	279	15	US-10-106-698-5351 Sequence 5351, Ap
30	186.5	2.8	470	10	US-09-755-830-13 Sequence 957, App
31	186.5	2.8	600	10	US-09-764-864-957 Sequence 957, App
32	188	2.7	1051	15	US-10-097-340-79 Sequence 79, Appl
33	187.5	2.7	1756	15	US-10-024-453-2 Sequence 2, Appl
34	187.5	2.7	1756	15	US-10-142-650-3 Sequence 3, Appl
35	187	2.7	532	10	US-09-801-368-350 Sequence 360, App
36	186.5	2.7	1258	15	US-10-198-070-101 Sequence 101, App
37	185	2.7	3067	12	US-09-949-029-18 Sequence 18, App
38	184.5	2.7	730	10	US-09-601-368-126 Sequence 126, App
39	182.5	2.7	549	10	US-09-801-368-358 Sequence 358, App
40	182.5	2.7	1506	15	US-10-232-561-12 Sequence 12, Appl
41	182	2.7	1127	15	US-10-232-561-12 Sequence 12, Appl
42	179.5	2.6	464	10	US-09-842-777-10 Sequence 10, App
43	179.5	2.6	1523	12	US-10-205-215-135 Sequence 135, App
44	176.5	2.6	1042	15	US-10-097-340-83 Sequence 83, Appl
45	174.5	2.5	1551	9	US-09-864-761-35904 Sequence 35904, A

#### ALIGNMENTS

#### RESULT :

```
US-10-177-293-472
; Sequence 472, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarpu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Heerssch, Sebastian
; APPLICANT: Xonahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT
; TIME OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/293,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
```

; PRIOR APPLICATION NUMBER: US 60/xxx,xxx  
 ; PRIOR FILING DATE: 2002-05-14  
 ; NUMBER OF SEQ ID NOS: 506  
 ; SOFTWARE: Fast-Seq for Windows Version 4.0  
 ; SEQ ID NO 472  
 ; LENGTH: 1281  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-177-293-472

Query Match 100.0%; Score 6851; DB 15; Length 1281;  
 Best Local Similarity 100.0%; Pred. No. C;  
 Matches 1281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MVRKNPPLRNVASEGQILEPIGTESKVGSKNKEFSADQKSENDDQSDAAELNKEEH 60
DB 1 MVRKNPPLRNVASEGQILEPIGTESKVGSKNKEFSADQKSENDDQSDAAELNKEEH 60

QY 61 SLHVODPSSSSKKDLKSAVLSEKAGFNYESPSKGNFPPSPHDEVTRNMLAFSPAAAG 120
DB 61 SLHVODPSSSSKKDLKSAVLSEKAGFNYESPSKGNFPPSPHDEVTRNMLAFSPAAAG 120

QY 121 VCEPLKSPQAEADDPDMACTSGDSLETKEQKMSPKATETGQAQSGQANCOGLSPV 180
DB 121 VCEPLKSPQAEADDPDMACTSGDSLETKEQKMSPKATETGQAQSGQANCOGLSPV 180

QY 181 SVASKNPQVPSDGGVRLNKSKTDLVNDNPDPAPLSPQLQDFKCNICGYGNGPTDLI 240
DB 181 SVASKNPQVPSDGGVRLNKSKTDLVNDNPDPAPLSPQLQDFKCNICGYGNGPTDLI 240

QY 241 KHRKXHYHLGHNRTRODAELDSKILAHNMVQSHSKDFOKVNRVPSGVLODINSSRPV 300
DB 241 KHRKXHYHLGHNRTRODAELDSKILAHNMVQSHSKDFOKVNRVPSGVLODINSSRPV 300

QY 301 LUNGTYDVQVTSGGTFIGIKRKYPCDCGNTKYFRCKFCNFTYMGNSSTLEQHFLOTHPN 360
DB 301 LUNGTYDVQVTSGGTFIGIKRKYPCDCGNTKYFRCKFCNFTYMGNSSTLEQHFLOTHPN 360

QY 361 KIKASLPSSSEVAKPSEKNKSKS:PALQSSDSGLKMQDKKITVKGADDTFVGYSPKIPK 420
DB 361 KIKASLPSSSEVAKPSEKNKSKS:PALQSSDSGLKMQDKKITVKGADDTFVGYSPKIPK 420

QY 421 DSSQNGTEATSYWCKPFCSPSESSSLKLEHYGKHCAVQSGINPELNDK:SRGSV 480
DB 421 DSSQNGTEATSYWCKPFCSPSESSSLKLEHYGKHCAVQSGINPELNDK:SRGSV 480

QY 481 INQNDLAKSSSEGETMTKTDKSSGAKKDFSSKGAEDNMVTSYNCCFCDFRYSKSGPDPV 540
DB 481 INQNDLAKSSSEGETMTKTDKSSGAKKDFSSKGAEDNMVTSYNCCFCDFRYSKSGPDPV 540

QY 541 IVWGFLRHYYOQLNHNKTIKHCPFCPRGLCSPEKHJGEITYPFACRKSNSCHCALLLL 600
DB 541 IVWGFLRHYYOQLNHNKTIKHCPFCPRGLCSPEKHJGEITYPFACRKSNSCHCALLLL 600

QY 601 HLSPCAAGSSRVKHQCHQCSFTTDPDVLVLFHYESVHESQASDVKQEAHNLQSDGQCV 660
DB 601 HLSPCAAGSSRVKHQCHQCSFTTDPDVLVLFHYESVHESQASDVKQEAHNLQSDGQCV 660

QY 661 KESKEHCTKCDFTQVVEEISIRHYRAHSCYKRCQCSFTTAADTQSLLEHFNTHVHCQED 720
DB 661 KESKEHCTKCDFTQVVEEISIRHYRAHSCYKRCQCSFTTAADTQSLLEHFNTHVHCQED 720

QY 721 ITTANGEDGHAISTIKEEPKIDFRVYNLLTPDSKMGEPVSVVREKLEEKDGLKEKV 780
DB 721 ITTANGEDGHAISTIKEEPKIDFRVYNLLTPDSKMGEPVSVVREKLEEKDGLKEKV 780

QY 781 WTESSDDLNRVTRGADILRGSPSYTQASLGLLTPVSGTQTKTLRDSPNVAAHLAR 840
DB 781 WTESSDDLNRVTRGADILRGSPSYTQASLGLLTPVSGTQTKTLRDSPNVAAHLAR 840

QY 841 PIYGLAVETKFLQAPAGGKSGALPQQVPASGENKSKDESOSLLRRRGSGVFCANCC 900
DB 841 PIYGLAVETKFLQAPAGGKSGALPQQVPASGENKSKDESOSLLRRRGSGVFCANCC 900
  
```

```

QY 901 TTNTSLWRKXANGGYVCNACG:YQKLHSTPRPLN:IKONNGECIIRRTTRKRLNPEALQA 960
DB 901 TTNTSLWRKXANGGYVCNACG:YQKLHSTPRPLN:IKONNGECIIRRTTRKRLNPEALQA 960

QY 961 EQLNKQCRGSNEFQVNGSPLEPRSEDHLTESHQREIP:PLPSLSKYEAQSLTKSHSAQOPV 1020
DB 961 EQLNKQCRGSNEFQVNGSPLEPRSEDHLTESHQREIP:PLPSLSKYEAQSLTKSHSAQOPV 1020

QY 1021 LVSQTLDIHKRMQPLHIQ:KSPQESTGPGKSSSVSECKSGSERGSP:IEKYVRAKHPNY 1080
DB 1021 LVSQTLDIHKRMQPLHIQ:KSPQESTGPGKSSSVSECKSGSERGSP:IEKYVRAKHPNY 1080

QY 1081 SPGSPDIEKYQVYELFC:PFVHNDPSEADMLRFWSKYKLSVPGNPHYLSHVPGLNPNQCN 1140
DB 1081 SPGSPDIEKYQVYELFC:PFVHNDPSEADMLRFWSKYKLSVPGNPHYLSHVPGLNPNQCN 1140

QY 1141 YVPYPTENLPHFSAVGSDND:PLDLA:KHSRPGPTANGASKEKTKAPPNVKNEGP:LVV 1200
DB 1141 YVPYPTENLPHFSAVGSDND:PLDLA:KHSRPGPTANGASKEKTKAPPNVKNEGP:LVV 1200

QY 1201 KTEKVDKSTODELS:KCVHCGTYVFLDEVMYALHMSCHGDSGPFQCS:CHLCTDKYDETT 1260
DB 1201 KTEKVDKSTODELS:KCVHCGTYVFLDEVMYALHMSCHGDSGPFQCS:CHLCTDKYDETT 1260

QY 1261 HIQRLRNNAQVEKNKPKPE 1281
DB 1261 HIQRLRNNAQVEKNKPKPE 1281
  
```

## RESULT 2

US-10-037-667-1  
 ; Sequence 1, Application US/10017667  
 ; Publication No. JG200202:77145A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Motcar, Bruce A.  
 ; TITLE OF INVENTION: REGULATION OF NEURAL DEVELOPMENT BY  
 ; TITLE OF INVENTION: LADDAJOS  
 ; FILE REFERENCE: 10287-04400  
 ; CURRENT APPLICATION NUMBER: JG/10/037,667  
 ; CURRENT FILING DATE: 2002-07-23  
 ; PRIOR APPLICATION NUMBERS: 60/243,110  
 ; PRIOR FILING DATE: 2000-10-25  
 ; NUMBER OF SEQ ID NOS: 13  
 ; SOFTWARE: Fast-Seq for Windows Version 4.0  
 ; SEQ ID NO 1  
 ; LENGTH: 537  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 ; US-10-037-667-1

Query Match 3.6%; Score 246; DB 14; Length 537;  
 Best Local Similarity 19.7%; Pred. No. 2,9e-10;  
 Matches 151; Conservative 82; Mismatches 235; Indels 338; Gaps 32;

```

QY 498 TDKSSGAKKDFSSKGAEDNMVTSYNCCFCDFRYSKSGPDPV 548
DB 29 TPNSQSSPSRSJANSIKVEM:-----YSDESSRLGPDRLDLDKDDSVVDEDSUSE 81

QY 549 HYQQLNHNKTIKHCPFCPRGLC:--SPERHLGITYPFACRKSNSCHCALLLLHLSFG 605
DB 82 -----PLGYCDGSGGEPH-----SPG 97

QY 606 AAGSSRVKHQCHQCSFTTDPDVLVLFHYES:-----VHESQ:ASDVKQEAHNLQSD 655
DB 98 G:ELPNGKUKCDVCGXVCIGPVLVYHVKESH7GESPPHCNQCASFTQKGLRLHKLHLS 157

QY 656 GQGSVKESKHSCTKCDFTQVVEEISIRHYRAHSC:-----YKRCQCSFTACTOSL 707
DB 158 GEKPFK:-----CPFCNYACRRDALTGHL-RTHSVSSFTVCKPKYCNVCGRSYKQOSTL 210

QY 708 LEHNTVTHCQED:ITTANGEDGHAISTIKEEPKIDFRVYNLLTPDSKMGEPVSVVYK 767
DB 708 LEHNTVTHCQED:ITTANGEDGHAISTIKEEPKIDFRVYNLLTPDSKMGEPVSVVYK 767
  
```

```

Db 211 EEKCHNYQLSLST-----DAQALT-----CQP-----235
QY 768 EKLEKGLKVKWTESSDOLRVNWRGADILRGS---PSYQASLGLTTPSGTQECT 824
Db 236 -----GDBIRLEWVPSMLHPSTERTPTFIDRLANSUT-----268
QY 825 KTLRDSNVAAHLARPIYGLAVETKGFLOQAPAGGKSGALPQQVPASGENKSKDESOS 884
Db 269 -----KREKSTPKF---VGEKQMRFLSD 290
QY 885 LRRRRGSGVFCANCLTTKSLMRKNAGVVCNACGJYQKLHSTPLNIIKQNGEQZ 944
Db 291 -----PYDVNASGGY-----EKL 304
QY 945 IRRTRKRLNPEALQAEQLNQQRSGNEEQVNGSPLERRSEDLHTEHQREIPLPLSLKY 1004
Db 305 VELVAHGLEPFG-----GSLAFVGTGHL-----APLRLP?-----336
QY 1005 EAQSLTKSHSAQOVLVVSOTLDIHKRMQPLHIQIKSPQESTGCPGNSSSVSEKGSSE 1064
Db 337 -----TNCISELTPVISS---VVTQMPTPSRLPLP-----GSRAGEGPELDGD 378
QY 1065 GSPTEKYNRPKHNYPSPGSPIEKYQYPLPGLPFVENDQSEADWLRFSKYKLVPGN 1124
Db 379 GGPL---LYRAGSLTDGASP-----SNGQDSTC-----TESKH 410
QY 1125 PHYLSHVPLEN---PCQNVYVPTFNPPHPSAVGSDNDIPDLAIKHSRPGPTANGASK 1182
Db 412 EDRIQGVVSLQGGPPQP---PPPTIVGRHSPAVAKEDPKPQE-GLLRGTGP?-----SK 462
QY 1183 EKTAPPNVKNEGLPNVVKTEKVRSTODELSTKCVHCGIVFLDEVYALHMSCHGDSGP 1242
Db 463 EVLRV-----VGESGPVKAFKCEHCRIFLDHVMTTHVCHGCHFRP 505
QY 1243 FOCSCQHLCTDKYDFTTHIORGLHR 1268
Db 506 FECNICGHSQDRYEFSSHIVRGEK 531

RESULT 3
US-10-037-667-4
; Sequence 4, Application US/10037667
; Publication No. US20020177145A1
; GENERAL INFORMATION:
; APPLICANT: Morgan, Bruce A.
; TITLE OF INVENTION: REGULATION OF NEURAL DEVELOPMENT BY
; FILE REFERENCE: 10287-044001
; CURRENT APPLICATION NUMBER: US/10/037,667
; PRIOR FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: 60/243,110
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-037-667-4

Query Match 3.5%; Score 242; DB 14; Length 515;
Best Local Similarity 23.6%; Pred. No. 5,6e-10;
Matches 127; Conservative 66; Mismatches 199; Indels 146; Gaps 25;

QY 814 LTPVSTQEQTKLTD-SNVVAAHLARPIYGLAVETKGFLOGAPA---GGEKSGALPQQ 869
Db 37 LSTTSCAQNSKSDRGMSNVK-----VETQSDENGRCACMGEECAEDLRV 84
QY 870 YPASGENKS---KDESQSLRRRRSGV-----FCA-NCLTTKTLMRKNAN 912
Db 85 LQASGEMNGSHRDQSSAL-----SGVGIRLPLNGKUKDCDIGVICVGPVLMVHRKSH 139
QY 913 GG---YVCNACGL-----YQKLHSTPLNIIKQNGEQIIRRTKRLNPEALQ 959

```

```

Db 140 TGERPFOCNOGASFTQGNLRLRHILKHSKEP---FKCHLCNVACRRR-----DALT 189
QY 960 AEQANKQQRGSNEQVNGVSPLEERS--EDHLTSHQ--REIPLPSL-----SKYEAQ 1007
Db 130 GHJRTSHVKGFKKCGYCGRSYKQKSSLEEHKEKCHNYLESMLGFCVCPV:KEETNNKEMA 249
QY 1008 GSLTKSHSAQCPV:VSQTLDIHKRMQPLHIQIKSPQESTGCPGNSSSVSEKGSSESGSP 1367
Db 250 EDLCIKIAGERSLVLDRLASNVAKKSSM-----PQKFJGCKCLSDMPYDSANVEDMM 303
QY 1068 IEKYNRPKHP--NY-----SPQSS-----PTEKYQY-----PLGLFVNDPQ 1105
Db 304 TSHYDCAINNAIYGAESLRLPVQTPGSSSEVVPVISSYQLHKPPSDGPPSRHSAQ 363
QY 1106 SEADWLRFSKYKSLVPGNPHYLSHVPLP-NPCQNVVP-----YPTNCLRP 1151
Db 364 DAVNLLLSKAK-SVS-----SERASPSNCCQDSTUTESNAEORSGLIYLTHNINP 416
QY 1152 HPSAVGSDNDIPDLAIKHSRPGPTANGASKETKAPPRYKNEGPLNVKVKTEKVDSTQD 1212
Db 412 R-----ARNGALJKEBQRAVEVLRAASENSQCAFVWSTSGE 453
QY 1212 ELST-KCVHCGIVFLDEVYALHMSCHGDSGPQCS-CQHLCTDKYDFTTHIORGLHR 1268
Db 454 QLVYKCECHVFLDHVMTTHVCHGCHFRDPFECNMGYHSDRYEFSSHIRGEHR 512

RESULT 4
US-10-205-823-321
; Sequence 321, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbatcheva, Bella
; APPLICANT: Hoerssch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Wonsley, Angela Y.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; PRIOR FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 321
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-205-823-321

Query Match 3.5%; Score 240; DB 15; Length 419;
Best Local Similarity 25.7%; Pred. No. 5.7e-10;
Matches 119; Conservative 47; Mismatches 167; Indels 130; Gaps 25;

QY 853 LOGAPAGGKSGALPQQVPASGENKSKDESQSLRRRPGSGVFCANCLTTKTLMRKNAN 912

```

[illegible]

RESULT 5  
CS-09-755-830-38  
; Sequence 38, Application US/09755830  
; Patent No. US2002010412A1  
; GENERAL INFORMATION:  
; APPLICANT: Georgopoulos, Kalia  
; TITLE OF INVENTION: IKAROS REGULATORY ELEMENTS AND USES

Query Match	3.5%	Score 238;	DB 10;
Best Local Similarity	25.8%	Pred. No. 9.6e-10;	
Matches 128;	Conservative	58;	Mismatches 194;
		Indels	116;
		Gaps	25;
		Length 461;	

```

119 KCHLCNYACRRDAATGHHJRTHSVGRKPGCGYGRSYKQRTSLEEH-----KERCHNYJ 172
QY
389 TESHQREIPLSLSKYEAQGS-----LTKSHSAQQPVLVQSQTLDJHKRMQCPHQIKSQP 1043
Db
173 ESXGLPGLYLPVKEETHSEMAEDLCJISERSVLDRLASINAKRKSSM-----PQ 225
QY
1044 ESTGDPNS-----SSVSEGGKSSERGSPLEKYMPPAKHPNV-----SPGGS 1085
Db
226 KFLGQXGLSDTPYSATYKEKENEMKSHVQDAINNA--INV-GABSLAPLVQTPPGSGE 283
QY
1086 --PIEKYQYALF---GLPFWNDQOSEA-DMLRFSWKYKLSVPGNPHYLSHVGLP-LP 1137
Db
294 WVEVISPMTQLRSEGTSPRSNAQDSAVEYLLLSKAL-VP-----SREASPSNS 336
QY
1138 CQKYVPVPTFNPPHSAVGSNDIP----LDLAIKHSRPGPTANGASKKTKAPPNVK 1192
Db
337 CQDSTTESNEQRSLYLTNHIARRAQRVSLKEH-RAVDLLRAASENSQDA----- 390
QY
1191 NEGFLNVNTEKVDKSTQBELSTKCVHCGIVFLDEWVYALHMSCHGSGPFOCSIOHLC 1252
Db
391 ---LAWST---SGEQMKVYKCEHRCRVFLDHWYTIHXGCHGFROPFCNMCGYHS 441
QY
1253 TDKYDTTHIQRLHR 1268
Db
442 QERYEFSSHTTGEHR 457

```

```

RESULT 6
US-09-755-830-40
; Sequence 40, Application US/09755832
; Patent No. US2020104112A1
; GENERAL INFORMATION:
; APPLICANT: Georgopoulos, Karia
; TITLE OF INVENTION: IKAROS REGULATORY ELEMENTS AND USES
; FILE REFERENCE: 10287-067001
; CURRENT APPLICATION NUMBER: US/097555.830
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 08/283,330
; PRIOR FILING DATE: 1994-07-29
; PRICE APPLICATION NUMBER: US 05/235,212
; PRIOR FILING DATE: 1994-05-02
; PRIOR APPLICATION NUMBER: US 08/121,438
; PRIOR FILING DATE: 1993-09-14
; PRIOR APPLICATION NUMBER: US 07/946,233
; PRIOR FILING DATE: 1992-09-14
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Fas-SEQ for Windows Version 4.0
; SEQ ID NO 40
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-755-830-40

```

Query Match	3.4%	Score 233.5	DB 20	Length 518
Best Local Similarity	23.3%	Pred. No. 2.6e-59		
Matches 126	Conservative	70	Mismatches 192	Indels 157
				Gaps 25

```
QY 1004 YEAGSLTKSHAQOVLVSOTLDIHKRMOPLIHQIKSPQESTGDPGNSSSVSEKGSSE 1063
Db 246 NEMAEEDLCKIGAERSLVLDRLASNAVKRKSMM-----PKFJGDKLSDMPYDSANYEK 239
QY 1064 RGSPIEKYMRPAKIP--NY-----SPPOS-----PIEKYCY-----PLFGLPFVH 1101
Db 300 EDMTSHYMDQAINNAINYLGAESLRPLVOTPPGSSEVVPVISSMYQLHKPPSDGPPRSN 359
QY 1102 NDFQSEADWLRFWSKYKLSVPGNPHYLSHVPLGP--NPCQNVYP-----YPTF 1147
Db 360 HSAQDAVDNLLLSAK--SVS-----SERASPSNSCQDSTDESNAEQRSLIYLTN 412
QY 1148 NLPPHFSVAGSDNIDPLDLAIKHSPGPTANGASKETKAPPNVKNBGPVNVKTEXVDR 1207
Db 413 HINPH-----ARGLAKEQRAYEVLRAASENSQDAFEVVS 449
QY 1208 STQDELST--KVCHGIVELDEVWYALHMSCHGDSG---PPOCSICQHLCTCKYDFTTHIQ 1263
Db 450 TSGEQLKYKCHCRVLFDLHVMTYIHMCHGCHGFRDPFECNMGYHSQDRIYFSSHIT 509
QY 1264 RGLHR 1268
Db 510 RGEHR 514

RESULT 7
US-09-755-830-43
; Sequence 43, Application US/09755830
; Patent No. US20020104112A1
; GENERAL INFORMATION:
; APPLICANT: Georgopoulos, Katia
; TITLE OF INVENTION: IKAROS REGULATORY ELEMENTS AND USES
; FILE OF INVENTION: THEREOF
; FILE REFERENCE: 10287-067001
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 08/283,300
; PRIOR FILING DATE: 1994-07-29
; PRIOR FILING DATE: 1994-05-02
; PRIOR FILING DATE: 1991-09-14
; PRIOR FILING DATE: 1992-09-14
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 334
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-755-830-43

Query Match 3.3%; Score 225; DB 10; Length 334;
Best Local Similarity 24.2%; Pred. No. 5.8e-09;
Matches 94; Conservative 53; Mismatches 145; Indels 96; Gaps 17;

QY 915 YVCNACGL-----YQKLHSTPRPLNIKON-----NGEQLIRRTRK 951
Db 5 PQCNQCGASFTQKGNLLRHILKHSCEKPKCHLCNYACRRDALTGHURTHSVIKETK- 63
QY 952 RUNPEALQAEQNLK--QQRGSNEEOVNGSPLEERS---EDULTESHQREIPLPLSLKYEA 1006
Db 64 ----HSEWAEEDLCKIGSERSLVLDRLASNAVKRKSMMPKCFGDKGLSDTPYDS-ATYEK 1118
QY 1007 QGSLTKSHAQOVLVSOTLDIHKRMOPLIHQIKSPQESTGDPGNSSSVSEKGSSE 1066
Db 119 ENEMMKSHVMDQAINNAINYLGAESLRPL--VQTP-----PG-----GS 105
QY 1067 PIEKYVRRPAKHNYSPGSPPIEKYQYPLFGLPFVHNDPQSEADWLRFWSKYKLSVPGNPH 1126
Db 156 EVVPVISPWYQVHRRSEGTSPRSN-----HSAQDSAVEYLLLSAKL--VP----- 1199
```

```
QY 1127 YLSHVPLGP--NPCQNVYP--FNLPHPHFSVAGSDNDIP-----LDLAIKHRRPQPTANGA 1180
Db 200 --SEREASPSNSCQDSTDESNAEQRSLIYLTHIARRACRVSJLKEEH-RAYDLLRAA 256
QY 1181 SKEKTKAPPNVKNBGPVNVKTEKVDKSTQDELSTCKVCHGIVELDEVWYALHMSCHGDS 1240
Db 257 SENSODA-----LRVST-----SGEQMYKCHCRVLFDLHVMTYIHMCHGFR 302
QY 1241 GPQCSICQHLCTCKYDFTTHIQRLHR 1268
Db 303 DPFEKNMGYHSQRIYFSSHITRGEHR 330

RESULT 8
US-09-755-830-39
; Sequence 39, Application US/09755830
; Patent No. US20020104112A1
; GENERAL INFORMATION:
; APPLICANT: Georgopoulos, Katia
; TITLE OF INVENTION: IKAROS REGULATORY ELEMENTS AND USES
; FILE OF INVENTION: THEREOF
; FILE REFERENCE: 10287-067001
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 08/283,300
; PRIOR FILING DATE: 1994-07-29
; PRIOR APPLICATION NUMBER: US 08/216,212
; PRIOR FILING DATE: 1994-05-02
; PRIOR FILING DATE: 1993-09-14
; PRIOR APPLICATION NUMBER: US 08/216,438
; PRIOR APPLICATION NUMBER: US 07/946,233
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-755-830-39

Query Match 3.3%; Score 224.5; DB 10; Length 432;
Best Local Similarity 22.3%; Pred. No. 9.7e-09;
Matches 114; Conservative 66; Mismatches 156; Indels 175; Gaps 22;

QY 814 LTPVSGTQEQTKLURD--SPNVAAHLARPLYGLAVETKGFLOGAPA---GGEKSGALPQQ 869
Db 37 LSTTSGAQONSKSDRGVSNVK-----VETQSDENGRCACEMNGEECAEDLRM 84
QY 870 YPASGENKS---KDESQSLLRRRSGGVFCANCCTTKTSLARKVANGYVNCACGL---- 922
Db 85 LDASGEKNGSHRDGSSAL-----SGVGG-----RLPNGKLKCDICGIVCIG 128
QY 923 -----YQKLHSTPRP-----LNIKONKGEQ-----I-RRRTTRKRLNP 955
Db 129 PNVLVYHRSHTGERPFQCNCGASFTQKGNLLRHILKHSCEKPKCHLCNYACRRR--- 185
QY 956 BALQAEQNLKQQRGSNEEOVNGSPLEERSDEHLTESHQREIPLPLSLKYAQSGLTSKHS 1015
Db 186 -----DALTGHLRTHSGDKCLSDMPYDS-ANYEKEDMMT-SHV 221
QY 1016 ACPQVLVSCOTLDIHKRMOPLIHQIKSPQESTGDPGNSSSVSEKGSSESPLEKYMRPA 1075
Db 222 KDOAINNAINYLGAESLRPL--VQTP-----PGSEVVPVISSMYQLH 262
QY 1076 KHPNYSPPSPPIEKYQYPLFGLPFVHNDPQSEADWLRFWSKYKLSVPGNPHY-SHVPLGP 1125
Db 263 KPESSGPPRS-----HSAQDAVDNLLLSAK-SVS-----SEREASP 300
QY 1136 -NPCQNVYP-----YPTFNLPPHFSVAGSDNDIPDLAIKHRRPQPTANGAS 1181
Db 301 SNSCQDSTDESNAEQRSLIYLTHIINH-----ARGLA 337
QY 1182 KEKTKAPPNVKNBGPVNVKTEKVDKSTQDELST--KVCHGIVELDEVWYALHMSCHGDS 1240
```

```

338 LKGEQRAYEVLRAASENSQDAFRVWSTSGQLKVKCEHCRVLFLDHWYTHYGHSC 397
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1241 G---PFOCSTCOHLCXTDKYDFTHIQRLHR 1268
      -:::--:--:--:--:--:--:--:--:--:--:--:--:--:--:--:--:--
398 GFROPFECNMGYHSODRYEFSHITRGEHR 428
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

RESULT 9
US-09-764-864-1095
; Sequence 947, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 947
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-764-864-1095

Query Match      3.2%; Score 222; DB 10; Length 323;
Best Local Similarity 27.2%; Pred. No. 9.5e-09;
Matches 97; Conservative 38; Mismatches 128; Indels 94; Gaps 20;

QY 367 QRGSEQVNGSPLERRSECHLTESHQREIPLPSLSKYEAQSLTKSHSAQOPV---LVS 1023
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 8 ETGEKPHRCHLCPFASAYERHL-EAHR 48
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

QY 1024 -----QTLDIHKRMQPLHIQKSPQESTGDPGNSSSVSEK--GSSERGSPTEKYMRA 1075
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 49 FRGSDRNLSHRRRKHKMVPK-----GTRSLSSKKWGVLCCKTKNLGYSRRA 99
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

QY 1076 KHPNYSPPGSPTEKYQYPLFGLPFVHNDFOSEADMLR---FMSKYKLSVPG----- 1123
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 100 L-INLSPPSMVWQKPDY-----LNDFTHEIPNICTDSYESMAKTTPTGGLPRCPOEL 150
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

QY 1124 ---NP-HYLSHVPG-----LNPQCN-----YVYPF---TFNL-PPHFSAVGS--DNDIPL 1164
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 151 MVQNPQLNLSTLAGQLSSLPENQNPASPDVVPDPEKPFMIQOPSTQAVWSAVSASIPQ 210
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

QY 1165 DLAKHSRPGPTAN-----GASKEKT--KAPPNVKNEGPLNVVVKTEKVDSTQD-EL 1213
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 211 SSPTSPEPRPSSHQRNYSVPAGSPSEPSAHTSTPSIGNSQP-----STPAPALPVQCDPQL 266
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

QY 1214 STKCVHCGIVFLDEVYALHMSCHSDSGPPQCSICQHLCTDKYDFTHIQRLHRNN 1270
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 267 LHKQCHQCDMYFADNLYTIHMGCHGYENPFQCNCGCKCKNKKYDFACHFARGQHCH 323
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

RESULT 10
US-09-764-864-1515
; Sequence 1515, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1515
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (84)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-09-764-864-1515

Query Match      3.2%; Score 217.5; DB 10; Length 294;
Best Local Similarity 28.5%; Pred. No. 1.8e-08;
Matches 39; Conservative 35; Mismatches 113; Indels 75; Gaps 18;

QY 1012 KSHSAQOPV---LVS-----QTLDIHKRMQPLHIQKSPQESTGDPGNSSSVSEK--G 1060
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 5 RSHTGEKPYKCELCSFRCSDRNLSHRRRKHKMVPK-----GTRSLSSKKWVG 55
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

QY 1061 SSERGSPTEKYQYPLFGLPFVHNDFOSEADMLR---FMSKY 1117
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 56 VJOKKTSNLGYSRRA---LNLSPSMVWQKPDY-----LNDFTHEIPNICTDSYESMA 106
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

QY 1118 KLSVPG-----NP-HYLSHVPG-----LNPQCN-----YVYPF---TFNL-PP 1151
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 107 KTTPTGGLPRDQELMVENPQLNLSTLAGQLSSLPENQNPASPDVVPDPEKPFMIQOP 166
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

QY 1152 HFSAVGS--DNDIPLDLAKHSRPGPTAN-----GASKEKT--KAPPNVKNEGPLNV 1199
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 167 STQAVWSAVSASIPQSSSTPSPEPRPSSHQRNYSVPAGSPSEPSAHTSTPSIGNSQP--- 223
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

```

```

338 LKGEQRAYEVLRAASENSQDAFRVWSTSGQLKVKCEHCRVLFLDHWYTHYGHSC 397
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1241 G---PFOCSTCOHLCXTDKYDFTHIQRLHR 1268
      -:::--:--:--:--:--:--:--:--:--:--:--:--:--:--:~:~:~
398 GFROPFECNMGYHSODRYEFSHITRGEHR 428
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

RESULT 9
US-09-764-864-947
; Sequence 947, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 947
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-764-864-947

Query Match      3.2%; Score 222; DB 10; Length 323;
Best Local Similarity 27.2%; Pred. No. 9.5e-09;
Matches 97; Conservative 38; Mismatches 128; Indels 94; Gaps 20;

QY 367 QRGSEQVNGSPLERRSECHLTESHQREIPLPSLSKYEAQSLTKSHSAQOPV---LVS 1023
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 8 ETGEKPHRCHLCPFASAYERHL-EAHR 48
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

QY 1024 -----QTLDIHKRMQPLHIQKSPQESTGDPGNSSSVSEK--GSSERGSPTEKYMRA 1075
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 49 FRGSDRNLSHRRRKHKMVPK-----GTRSLSSKKWGVLCCKTKNLGYSRRA 99
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

QY 1076 KHPNYSPPGSPTEKYQYPLFGLPFVHNDFOSEADMLR---FMSKYKLSVPG----- 1123
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 100 L-INLSPPSMVWQKPDY-----LNDFTHEIPNICTDSYESMAKTTPTGGLPRCPOEL 150
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

QY 1124 ---NP-HYLSHVPG-----LNPQCN-----YVYPF---TFNL-PPHFSAVGS--DNDIPL 1164
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 151 MVQNPQLNLSTLAGQLSSLPENQNPASPDVVPDPEKPFMIQOPSTQAVWSAVSASIPQ 210
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

QY 1165 DLAKHSRPGPTAN-----GASKEKT--KAPPNVKNEGPLNVVVKTEKVDSTQD-EL 1213
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 211 SSPTSPEPRPSSHQRNYSVPAGSPSEPSAHTSTPSIGNSQP-----STPAPALPVQCDPQL 266
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

QY 1214 STKCVHCGIVFLDEVYALHMSCHSDSGPPQCSICQHLCTDKYDFTHIQRLHRNN 1270
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 267 LHKQCHQCDMYFADNLYTIHMGCHGYENPFQCNCGCKCKNKKYDFACHFARGQHCH 323
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

RESULT 10
US-09-764-864-1095
; Sequence 1095, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1095
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-764-864-1095

```

QY 1200 VKTEKYDRSTQD-ELSTKCVHCOIVFLDEVYALHMSCHDSGPFQCSICQHLCTDKYDF 1258  
Dd 224 -STPAPALVQDPQLLHCHQCDMYFADNILYTIHMGCHGYENPQCNLCGCKCKNKYDF 282  
QY 1259 TTHIQEGLHNN 1270  
Dd 283 ACHFARGQHQN 294

RESULT 12

US-09-749-728B-11  
; Sequence 11, Application US/09749728B  
; Patent No. US20020142457A1  
; GENERAL INFORMATION:  
; APPLICANT: Umezawa, Akihiro  
; APPLICANT: Hata, Jun-ichi  
; APPLICANT: Fukuda, Keiichi  
; APPLICANT: Ogawa, Satoshi  
; APPLICANT: Sakurada, Kazuhiro  
; APPLICANT: Gojo, Satoshi  
; APPLICANT: Yamada, Yoji  
; TITLE OF INVENTION: THE CELL HAVING THE POTENTIALITY OF DIFFERENTIATION INTO CARDIOMY  
; FILE REFERENCE: 03746 000043  
; CURRENT APPLICATION NUMBER: US/09/749,728B  
; CURRENT FILING DATE: 2001-09-17  
; PRIOR APPLICATION NUMBER: H11-372826  
; PRIOR FILING DATE: 1999-12-28  
; PRIOR APPLICATION NUMBER: PCT-JP00-01148  
; PRIOR FILING DATE: 2000-02-28  
; PRIOR APPLICATION NUMBER: PCT-JP00-07741  
; PRIOR FILING DATE: 2000-11-02  
; NUMBER OF SEQ ID NOS: 80  
; SOFTWARE: PatentIn Ver.2.0  
; SEQ ID NO 11  
; LENGTH: 442  
; TYPE: PRP  
; ORGANISM: Homo sapiens  
US-09-749-728B-11

Query Match 3.1%; Score 215; DB 10; Length 442;  
Best Local Similarity 33.0%; Pred. No. 5.6e-08;  
Matches 69; Conservative 22; Mismatches 68; Indels 50; Gaps 9;  
QY 873 SGENKSKDEQSLLRRRRSGVFCANCLTKTSLMRNANGGVVNCAGLYOKLHSTRP 932  
Dd 248 NGINRELKPPQRLSARRVGLSCANCQTITTLRRNAEGEPVNCAGLYMKLHGVPR 307  
QY 933 LNI:KQNGEQ:IRRRTRKRLNFEALQAEOLNKQOR- - - - -GSEEQVNSPLRRSEHL 988  
Dd 308 LAWRKEG- - - - -IOTRKRKPKN- - - - -LNKSKTPAAPSGESLPPASGASSNSNAT 354  
QY 989 TESHQREIPL- - - - -PSLSKYBAQSLTKSHSAQPVLSQTL- - - - -DIHKEMQPL 1035  
Dd 355 TSSSEEMRPKTEPGLSSH- - - - -YGRSS- - - - -VSOTFSVSAMSGHPSIHPVLSAL 403  
QY 1036 HIQIKPOESTGDPGNSSSVSEKGSER 1064  
Dd 404 KL- - - - -SPQ- - - - -GYASPVSPQPSQ-SSK 423

RESULT 13

US-09-974-298-114  
; Sequence 114, Application US/09974298  
; Patent No. US20020156263A1  
; GENERAL INFORMATION:  
; APPLICANT: Chen, Hui-Mei  
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER  
; FILE REFERENCE: PA-0037 P  
; CURRENT APPLICATION NUMBER: US/09/974,298  
; CURRENT FILING DATE: 2001-10-04  
; PRIOR APPLICATION NUMBER: 60/238,331  
; PRIOR FILING DATE: 2000-05-10

; NUMBER OF SEQ ID NOS: 134  
; SOFTWARE: PERL Program  
; SEQ ID NO 114  
; LENGTH: 1125  
; TYPE: PRP  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. US20020156263A1 3267030CDD  
US-09-974-298-114

Query Match 3.1%; Score 211.5; DB 10; Length 1125;  
Best Local Similarity 13.8%; Pred. No. 4.9e-07;  
Matches 211; Conservative 138; Mismatches 377; Indels 411; Gaps 54;

QY 372 AKPSEKH-SNKSIPALQSSDGLGKW- - - - -QCKITVKG- - - - -LCTPVGVSVPRKL 420  
Dd 13 ANPRNNVNNYVTVETNSDSDDEKLHIVEESVTDADCEGVFEDDLPTQTV- - - - -LFG 70  
QY 421 DSSPQNGTEATSYWCKFCFSCESSSSJLLEHYKQKHGAVOSGLNPNELNKLRSV 480  
Dd 71 ESSEREENAKVCE- - - - -SDTSKEGCEI- - - - -LGPE- - - - - 98  
QY 481 INQNDIAKSSSEGTMTKTKSSSCAKKXDFSSKGAEDNXYTSVNCQFCDFRYSKSHGPDV 540  
Dd 99 - - - - -AAQDEAGCTYKDECESDAENQNHDPNVEFL- - - - -QQQDT 136  
QY 541 VVGPLLRHYQOLNTHKCT- - - - -IKHCPFCPRG- - - - -LCSPEKHLGSEITY 584  
Dd 137 AVIFPEAPEDCQGTPEASCHDENGTPJAFSOLLTCPCYCDRGYKFTSLKEH- - - - -IKYR 193  
QY 585 FACRKNCSHCALLLHLSPGANGSRVHKQCQCSTTPDVCVLLPHYESVHESQASDV 644  
Dd 194 HEKNEFNFS- - - - -CSLCSYT- - - - -FAVRTGLE- - - - - 217  
QY 545 KOEANHLOGSGQGSYKES- - - - -KEHSCTKCFPTQVEESISRYARAH- - - - -CYKCEQS 698  
Dd 218 RHMTSHKSGRC-QRHVTQSGCNRAFKTECGKAFKYHHLKEHL-RHSGEKPYECPNCK 275  
QY 699 FTAADTQSLLEHFNTHVQCEQDITTTANGEEDGHAISTIKEEPKIDFRVYNLLTP- - - - -DSXV 756  
Dd 276 KRFSHSGSYSSHSSKKC- - - - -ISLIPVNG- - - - -RRTG:KTSQCSFPS-SASF 321  
QY 757 GEPVSESVWREKLEBKDGLKE- - - - -KWTE- - - - -SSDDLANTWRG 796  
Dd 322 GSPTRPQI- - - - -RQK:ENKP-LCEQLSVNQIKTEPVDYEFKPIVVASG:NCSTPLQNGVFTG 378  
QY 797 ADILRGSPS- - - - -YTOA- - - - -SLGLTPVS- - - - -G 819  
Dd 379 GGPLQATSSPOGMVQAVVLTVLGVSP:SNLSQIONVLKVAVDGNVIRQVLENNQANLA 438  
QY 920 TQEQTKTLRDSPPVAAH- - - - -JAPYGLAVETKPFLOGAP- - - - - 857  
Dd 439 SKEQ-ETINASPIQQGSHSVISAISLPLVDQDQTTKIIINYSLEQSPSLQVWPQN:KKKN 497  
QY 958 - - - - -AGGEKSGALQCPASPGENSKDESQSLRRRSGSVFCANCI:TTXTSLMRKNANG 914  
Dd 498 PVATNSCKSEKGFEDLTVKSE- - - - -KDKSFE- - - - -GGVNSTCL- - - - - 533  
QY 915 YVCNACGLYQK:HSSTPPPLNI:KQNGEQ:IRRRTRKRLNFEALQAEOLNKQORGSNEEQ 974  
Dd 534 -LCDDC- - - - -EGDINALPELKHLD- - - - -KQTPQPPPLPAEAKEKPE-SSVSS 576  
QY 975 VNSGFLERRSECHTESHORE:PLPS-LSKYEAQSLTKSHSAQPVLSQCT- - - - -LD:HL 1029  
Dd 577 ATG- - - - -DGNLSFS- - - - -QPPLEKLLSLKAYVALNAQPSAEE:SK:ADSVNCPDVV 626  
QY 1030 K- - - - -RMQPLHIQ:KSPQESTGDPG- - - - -NSSSVSEK- - - - - 1059  
Dd 527 KAMFEKNQAQGIQVQSSSEPSPEPGKVNTPAKNDQFQSANANEPQDSTVNLQSPKMTN 686  
QY 1060 - - - - -SSSESGSFIEKYMRAKHP- - - - -NYSPPSPSIEKQYPL-FCLPFF 1100



```
Db 687 SPVLPVGTNGS---RSSTPSPPLNLSSRNTQCYLYTAEGAEEQVQFELDLSLPKQ 743
Qy 1101 HNDFOSEADWLRFWSKYKLSVPGNPHYLS-----HYVGLNPNQCN 1140
Db 744 QGELLERSTITSYQNSVYQVQEPNLNLSCAKKEPKQKSCVTDSEPVVNVIPPSANPINI 803
Qy 1141 VVPEYFENLPPHESAVGSDNDIPDLAIKHSRPGPTANGASKEKTKAPPNVK-----1192
Db 804 AITVTVAQPT-IVAIDQNSVPCLRAL-----AANKQITLIPQVAYTSTVTS 851
Qy 1193 ---NEGFLNVV-----KTEKVDNSTQ-----DELST-----KCVHCG 1222
Db 852 PAVQEPPLKVIQBNQDQERQDTSSEGVSNVEDQSDSTPPKKXMKRTENGMYACCLCD 911
Qy 1222 IVDLDEVMAVHMSCHGDSGPPQCSICQHLCTCKYDFTHIQRGLHRNNAQKVK 1278
Db 912 KIFQKSSLLRHKYETGRPHCEGICCKAFKHKKHLIEHMLHSGEKPYQCDKCK 968

RESULT 14
US-09-755-830-42
; Sequence 42, Application US/09755830
; Patent No. US20020104112A1
; GENERAL INFORMATION:
; APPLICANT: Georgopoulos, Kasia
; TITLE OF INVENTION: IKAROS REGULATORY ELEMENTS AND USES
; FILE REFERENCE: 10287-067001
; CURRENT APPLICATION NUMBER: US/C9/755,830
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 38/283,300
; PRIOR FILING DATE: 1994-07-29
; PRIOR APPLICATION NUMBER: US 38/238,212
; PRIOR FILING DATE: 1994-05-02
; PRIOR APPLICATION NUMBER: US 08/121,438
; PRIOR FILING DATE: 1993-09-14
; PRIOR APPLICATION NUMBER: US 07/946,233
; PRIOR FILING DATE: 1992-09-14
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-755-830-42

Query Match 3.1%; Score 210.5; DB 10; Length 376;
Best Local Similarity 22.3%; Pred. No. 9.6e-08;
Matches 107; Conservative 60; Mismatches 144; Indels 169; Gaps 21;

Qy 814 LTPVSGTQETKLRD-SPNVEAAHLARPIYGLAVETKGFJQCAPA---GGKSGALPQQ 969
Db 37 LSTTSGAQNSKDRGMASNVK-----VETQSDENGRAEENGECAEDLRM 84
Qy 870 YPASGENKS---KDESQSLRRRRGSGVFCANCLTKTSLMRKNVANGVYVQACGLYQKL 926
Db 85 LDASGEKMGNSHRDQGSAL---SGVGI-----RLPMGLKLCDCIGIV---125
Qy 927 HSTPRPLNIQKNNGEQIIRRTKRLNPEALQAEQNLKQORGSNEEQVNGSPLERRSDD 986
Db 126 -----CIGPNVLM-----VHKRS--138
Qy 987 HLTESHOREITPLSLSYEAGSLTKSHSAQOPVLVSQTLDIHKRQPHLIQIKPQEST 1046
Db 139 HTGDKCLSDMPYDS-ANYEKEDMMT-SHVMDQAINNAIYLGAEISURPL---VQTP-189
Qy 1047 GDPNGSSSVSEKSGSBERGSPFIEKYMPAPKPNYSPGSPFIEKYQVPLGLPFVHDFOS 1106
Db 190 -----PGSEVVPVSSMYQLHKPDSGDPRS-----NHSADQ 222
Qy 1107 EADWLRFWSKYKLSVPGNPHYLSVPLGP-NPCQNYVP-----YPTFNLPHP 1152
Db 223 AVDNLLLSKAK-SVS-----SREASPSNSQDQDSTTESNAEQRSGLYLTNHNHPH 275
```

```
Qy 1153 FSAVGSNDIPDLAIKHSRPGPTANGASKEKTKAPPNVKNEGPJNVVKTKEVDRSTQDE 1212
Db 276 -----ARNGLALKEBQRAYEVLAASENSGCCAFRVVSTSGEQ 312
Qy 1213 LGSF-KCVHCGIVFLDEVMAVHMSCHGDSG---PPQCSICQHLCTCKYDFTHIQRGLHR 1268
Db 313 LKVKYKCEHCVFLDHRVMTTHKGGCHGCHGRDPEFCNMCGYHSCDRYEFSSSHITRSEHR 372

RESULT 15
US-10-177-293-174
; Sequence 174, Application US/10-77293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Giatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarpu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Yareen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Roersch, Sebastian
; APPLICANT: Konahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: CONECTIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 174
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-293-174

Query Match 3.1%; Score 209; DB 15; Length 474;
Best Local Similarity 26.4%; Pred. No. 1.8e-07;
Matches 97; Conservative 33; Mismatches 107; Indels 130; Gaps 19;

Qy 750 LTPD-----SKMGEPVSESV-VKREKLEBKDLK-EKWTETESSDDJRNVTWKGADILRSP 804
Db 175 LSPDPTTGAASPASSAGSGARGEDKGVKYASLTESYKME-----SGRP 222
Qy 805 SYTQASLGLLTPVSGTQETKLRDSPNVEAAHLARPIYGLAVETKQ-----FLOCAP 857
Db 223 -----LRPGLAT--MGIQ-----PATHHPITPYPSVVPAAAHIDYSSGLFHPGSF 264
Qy 858 AGGEKSGALPQQ-----YPASGENK 877
Db 265 LGGPASSFTPKQREKTRSCSEGRECVNCGATATPLWRBDGTGHYICNACGFYHKMGQNR 324
```

```
QY      878 SKDESOLLRRRRSGVFCANCUTTTKTLWRKNVANGYVCNACGLYOKLHSTPRPLNKK 937
Db      325 PLIKPVERLSAARRAGTCCANCOTTTTLRRRNANGDPVCNACGLYKLNVRPEJTKK 384
QY      938 QNNGEQIIRRTKRLNPEALQAEQLNKOQGSN--EE-----QVNGSPLEERS-EDHLT 989
Db      385 EG-----IOTRNRKMN-----KSKKSKGAECFEELSKCMQEKSPFSAALAGHMA 432
QY      990 ESHQREIPLPSLSKYEAQGSLLTKSHSAQOPVLVSOTLD-HKMQPLHIQKSPQESTGDP 1049
Db      433 -----PMGHLPPF-----SHSGH--LEPTPT-----PIH---PSSLSFGHP 464
QY      1050 GNSSSVS 1056
Db      465 HPSSMVT 471
```

Search completed: October 29, 2003, 11:53:09  
Job time : 232 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 11:29:57 ; Search time 72 Seconds

(without alignments)  
1711.00i Million cell updates/sec

Title: US-09-702-216-2

Perfect score: 6851

Sequence: 1 MVRKNPPLRNVAEGECQI.....IQRGLHRNNAQVENKGFKE :281

Scoring table: BL0SUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_76:\*

1: Pir1:\*

2: Pir2:\*

3: Pir3:\*

4: Pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	271	4.0	1186	T33754	O/E-1-associated z
2	268	3.9	758	A29253	finger protein hun
3	248.5	3.6	1114	I50222	delta2 - chicken
4	241	3.5	452	C41602	transcription fact
5	239.5	3.5	518	A56355	DNA-binding protein
6	234	3.4	782	S04047	finger protein zfy
7	233	3.4	1350	S00647	finger protein - A
8	222.5	3.2	466	A36389	transcription fact
9	222.5	3.2	982	T43676	hunchback-related
10	218.5	3.2	816	S05548	gap protein hunchb
11	218	3.2	783	A31491	sex-determining re
12	215	3.1	439	I57561	transcription fact
13	215	3.1	480	A40815	transcription fact
14	213	3.1	392	A56229	lymphoid transcrip
15	213	3.1	779	S40382	box A-binding fact
16	212.5	3.1	440	I61183	transcription fact
17	212	3.1	431	I59572	Ikaros DNA binding
18	209.5	3.1	444	A39794	transcription fact
19	209	3.1	474	A41782	transcription fact
20	209	3.1	2783	A41948	alpha-fetoprotein
21	207.5	3.0	444	B36389	transcription fact
22	207	3.0	443	B39794	transcription fact
23	207	3.0	1124	JX0293	zinc finger protei
24	207	3.0	1828	A40115	microtubule-associ
25	206.5	3.0	390	I51419	transcription fact
26	206	3.0	427	B56229	lymphoid transcrip
27	205.5	3.0	439	A48099	transcription fact
28	205.5	3.0	1036	A34755	nitrogen regulator
29	205	3.0	380	I50701	transcription fact

transcription fact  
transcription fact  
GATA-type transcr  
transcription fact  
transcription fact  
eyeless, long form  
URSL protein - sm  
E-box-binding repr  
zfa protein - mous  
finger protein 2fy  
transcription fact  
transcription fact  
homeotic protein 2  
regulatory protein  
cell proliferation  
transcription fact  
microtubule-associ

## ALIGNMENTS

RESULT :  
T33754  
O/E-1-associated zinc finger protein Roaz - rat  
C/Species: Rattus norvegicus (Norway rat)  
C/Date: 02-Sep-2003 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000  
C/Accession: T33754  
R/Teal, R.V.L.; Reed, R.R.  
submitted to the EMBL Data Library, March 1997  
A/Description: Cloning and functional characterization of Roaz, a zinc finger protein t  
A/Reference number: Z21399  
A/Accession: T33754  
A/Status: Preliminary; translated from GB/EMBL/DBCB  
A/Molecule type: mRNA  
A/Residues: 1-1186 <TSA>  
A/Cross-references: EMBL:J52564; NID:G2149791; PID:G2149792; PIDN:AAB56646.;  
A/Experimental source: strain Sprague-Dawley  
C/Function:  
A/Description: regulates the temporal and spatial pattern of olfactory neuronal-specif

Query Match 4.0%; Score 271; DB 2; Length 1186;  
Best Local Similarity 20.0%; Pred No 1.3e-07;  
Matches 275; Conservative 156; Mismatches 515; Indels 432; Gaps 69;

QY	52	AELNHKEEHLHVQDPSSSSKKDLKSAVLSEKAGNYESFSGKGNPPSPPHDEVTDENKL	111
DB	84	AAPSRDHLKHLKTHSS--KPFKCSVC--KRGFSSTSLQS-----HVCARKKNK-	131
QY	112	AFSFPNAGGVCEPL-KSPQRAEDDPQDMACTPSGDSL-ETKEDQK-----VSPKATEETG	165
DB	132	-----EHLAKSEKAKKD---DFMCDYCEDTFSTCELEKHVJTLHPQLSEKAD	177
QY	166	-----QAQSGCANCOGLSPVSVVA-----SKNPQ	188
DB	178	LQCIHCPEVVDVESTLAHIOAHANQKHKPCWCEQFSVSEGVYCHLDSHRQDPSSNHS	237
QY	189	VPSD---GGVRLNKSTDLLVNDNPD-----APSPELQCFKCNICGYGYGNPDTDJ	240
DB	238	VSPDPVLGVSANSSATP---DSTPDVLGVSANSSATPDSSASV-----ERGSTPDSTL	290
QY	241	KHPRKVLGCHNRTQDAEJDSKILALHNMFVFSKQFQKYNRSVFGVGLQINKSRPV	300
DB	292	KPVRG-----QKMRDGGGWSKV---YSTPYCKRKDFTS---AVLEIHEKATINADRP-	339
QY	301	LANGTYDVQVTSGGTFIG---RKT-----PDQCGNTKYFRCKFCNFTVMGNSSTEL	350
DB	340	QSHTQICLDSMPTLYNLNENHVRKLKSHAPVWQFGNISAFHCNYCPKEMFADINS--J	397
QY	351	EQHFLOTH--PNKIKASLFSSEVAKPSEKXS-----NKSIPALQSSDSGDLGMQCKITVK	404
DB	398	QEHIRVSHCGPN-----ANPPDGNNAFFCNQCSMGFLTSS--LTEIIOAHCS	444
QY	405	AGD---DTPVGISVPIKPLDSSRQNGTEATSYVCKFCFSCESSSSSLKLLSHYKQHQGA	462

```
Db 445 VGSTKLESPV-----IQPTQSFME-----VSCPYCTNSPIFGSTLKLTKEHKN 492
QY 462 VOSGLPELNDKLSRGVINQNDLAKSSEGETMTKTKSSGAKKXDFSSKGAENMYT 521
Db 493 I-----PLAHSK-----KSKAQSPVSDVTEVSPKQRLGSA---NSIS 530
QY 522 S--YNCQFCDFRYKSHGPDVIVGPDLRHYQQLHNIHKCTIKHCPFCRGLCSPEXHLG 579
Db 531 NGEVPCNQCDLKPNFES-----FQTHLKLHLELLRKKQACPOCKEDFDSQSLQ 581
QY 580 EITVPFACRSN---CSHC-----ALLHLSPGAAGSSRVKHQCHQCSSTTPD 625
Db 582 HLTVHYVTTSTHYVCESCDKQFSSVDLQXHLDMH-----TFVYRCTLQCEVFD 633
QY 626 VDVLFFHVESVHESQAS-----DVKQEAHNLQSGDQSQVKES-----KEHSCTK 670
Db 634 KVSIOVLAVKHSNEKQMYRCTACNWDPRKEA-----DLQVHKVSHLGNPAKHKCIP 687
QY 671 CDFITQVEEISRYRAHSCYKRCQCSFTAAQTQSLEHFNTHVCOEQDITTTANGEEDG 730
Db 688 CGETFS-EVELOCHITTHSKYKNCRCFKAPFAVLLLEKHLREKHC-VFDEAAENGTA 746
QY 731 HAISTIKKEPKIDFRVYNLLTPDSKMGEPVSESVVREKLEEKDGLKEKVTWTESSCDL- 789
Db 747 VPP-STKRAEADLQGLMLKNPEA-----PNSH-----BASEDQVD 792
QY 790 RNTVMRGADILRGSPSYTOASLGLLTPVSGTQEQTKLRDSPNVVAAHARPIYGLAVET 849
Db 783 ASEPMYCDIC--GAATVMEV-----LQNHLELD-----HNRP----- 815
QY 850 KGFLQAPAGEKSGALPQQYPASGENKSKDEQSILRRRSGVFCANCLTTKTSLWRK 909
Db 816 -----GEDDGRKKAEFKGSHKCNVCSRTF-----PSENGLEHLQTHRG 856
QY 910 NANGYVCNAGLYQKLHSTPRPLNIKQNGEQIRRRPRKSLNPEALCAEOLNQQRG 969
Db 857 PAK-HYMCPICG-----ERFSLTLTEH-----KVTHSKSLD 886
QY 970 SNEQVNGSLPERSDEHLETSOREIPLPSLSKYEAQGSLL--KSHSAQPVLSQTL 1027
Db 889 TGTCTICKMPLQ--SEEFIEHCQW-----PDL-----RNSLTGFCVVCVGTVSTLEK 938
QY 1028 IHK--RMOPHLIQIKSQESTGDPGNSSSVSEKSGSERGPIEKYNRPKHPNYPSPGS 1085
Db 939 IHGTFHMQKL-----AGSSAAS-----SPNGQ 960
QY 1086 PIEKYQVPLFGLPFVHNDFOSEADMLRFWSKYLSVGNPHY-----SHVYGL- 1134
Db 961 GLQK-----LYKCALCLKEFRSKQDLVR-----LDVNGLPYGLCAGMARSANGVQGLA 1010
QY 1135 -PNPCQNVYPETNLPPHFSAVGSDNDIPDLAIKHSRPGPTANGASKETKAP-PNVK 1192
Db 1011 PPEPADR--PCAGLURCPENVCNKPESAEDLESHMQVCHRDLTPTETSGPRKCAQ-SPVPRK 1068
QY 1193 NEGLNVVVKTEKVDRTQ-----DELSLKVCHGIVFLDEVYALHM-----SCHGDS 1240
Db 1069 TYQCIKQMTFENEREQIHVANHMIIEGINHECKLCNQMFDPKALCLHLIEHSEFGMG 1128
QY 1241 GPFOCSIC-----OHL-----CTDKYDFTTHIQRLHNNNAQ 1272
Db 1129 GTFKCPVCTVFVQANKLQOHIFAVHQEDKIYDCSQCPKFFQTEQLQNHITSQRAQ 1186

RESULT 2
A29253
finger protein hunchback - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 07-Mar-1988 #sequence_revision 07-Mar-1988 #text_change 01-Dec-2000
C:Accession: A93395; A46363; A29253
R:Tautz, D.; Lehmann, R.; Schnuerch, H.; Schuh, R.; Seifert, E.; Kienlin, A.; Cores, K.;
Nature 327, 383-389, 1987
A:Title: Finger protein of novel structure encoded by hunchback, a second member of the
```

```
A:Reference number: A93395
A:Accession: A93395
A:Molecule type: DNA
A:Residues: 1-758 <TAU>
A:Cross-references: EMBL:Y00274; NID:98054; ZID:98055
R:Schmer, R.J.; Retzlaff, X.; Goerlich, K.; Sander, K.; Tautz, D.
Proc. Natl. Acad. Sci. U.S.A. 89, 10782-10786, 1992
A:Title: Evolutionary conservation pattern of zinc-finger domains of Drosophila segmen-
A:Reference number: A46363; MUD:93066327; PMID:1438276
A:Accession: A46363
A>Status: Preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 281-349 <SOM>
A>Note: sequence extracted from NCBI backbone (NCBIP:118691)
C:Genetics:
A:Gene: hunchback; FlyBase:hb
A:Cross-references: FlyBase:FBgn0001280
C:Keywords: DNA binding; nucleus; segmentation; zinc finger
```

```
Query Match 3.9%; Score 268; DB 2; Length 758;
Best Local Similarity 19.5%; Pred. No. 1e-07;
Matches 157; Conservative 98; Mismatches 301; Indels 248; Gaps 29;
```

```
QY 551 QCLHNTHKCTTKHCPFCRGLCSPEKH-----LGEITYPPACRKNCSHCALLLHLSFG 605
Db 115 QCHHHHHLXGGEKFLTPPG-LPNMQHFGYGNLRPSPTTTSASTAPVAV----- 166
QY 656 AAGSSRVKHQCHQCSFTTDDVDVLLFHYESVHESCASQKQEAJNELQSGDQSQVK---E 662
Db 167 ATGSS-----EKJQALTPPMDVT--PPKSPAKSQSNIEPEKEHDDQNSSEDMKYAE 218
QY 663 SKEH-----SCTKCDPIT 675
Db 219 SEDDDTNIPYIKSHGAKNKYKCTCGVA--TKVDFWHTRTMKPDKILQCKPCFPVT 278
QY 676 QVEEISHYR--RAHSCYKRCQCSFTAAOTQSLEHFN-----VHCQEQDITTT 723
Db 279 EFKHLEHYHIEKHKQKQFPQCDKCYTCVNKSMLSHRKSHSSVYQYRCACDQATKYCH 338
QY 724 ---ANGEEDGAI-STIKEE-----PKIDFRVYNLLT-PDSKMGSEFVSESVYKREKLEBK 774
Db 339 EFKLHLYKYGKGMVLDEDTGTPNPSLVICVGTGRGPKSNKGGF--ASS----- 387
QY 775 GLEKQWTESSDLENYTWGADILRGSPSYTOASLGLLTFVSGTQEQTKLRDSPNVE 834
Db 388 -----GSGSGS-----KSNVAAVAPQOQSQPA--QpVA-TSOLSAALQGFPLVQ 430
QY 835 ---ANFLARPIYGLAVETKGLQAPAGCEKSGALPQOYPASGENKSKDESQSLRRRG 891
Db 431 GNSAPPAASPVLP-----PASPAKSVASVEQTPSLPSP----- 464
QY 892 SGVFCANCLTTKTSLWRKNANGVYCNACG-YQKLHSTPRPLNIKQNGEQIRRRTRK 951
Db 465 -----ANLPPLASLQCNRMAPP-----PYNIN-----LQMLAAQQQNAV- 504
QY 952 RLNPALQAEQUNQCRGSENEEQNGVSPLERRSDHLCESHQREIPLPSLSKYBAQSLS 1011
Db 505 GLSPRYRQQLQQNQCCSDNQEEQDEYERKSVDSAMD------SQGTPV 550
QY 1012 KSHSAQPVLSQTLDIHKRMQPLHIQIKSQESTGDPGNSSSVSEKSGSERGSPIEKY 1071
Db 551 KEDEQQ-----QPOPLAMNLKVEEAT-PLMSSSNASRRG---RVKLDTL 595
QY 1072 MRPAKHNPYSP-----PGSPIEKYCPLFG---LPFVHNDFOSEADMLRFWSKYLSVPG 1123
Db 596 LQLRSEAVTSPEQLKVPSTPMPATASSPIAGRKPMPEEHCSGTSSAD----- 641
QY 1124 NPHYLSHVPGLPNPCQNVYPPTNLPPHFSAVGSDNDIPDLAIKHSRPGPTANGASK 1183
Db 642 ESMETAHVQANTSASSA-----SSSGNSNASSNS 673
QY 1184 KTKAPNKNKSGPLNVVVKYDRSTQDELSTKCHCGVFLDEVYALHMSCHGDSGPF 1243
```

Db 674 NGNSSSSSSNGTTSVAAPPSTPAAGAIYECKYCDIFFKDAVLVTHMGYSCDDVF 733  
QY 1244 OCSICCHLCTDKYDFTTHIQRLGH 1267  
Db 734 KNCMCEKCDGPGVLVTHMAVAH 757

## RESULT 3

I50222  
deltra5F1 - Chicken  
C:Species: Gallus gallus (chicken)  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 2i-Jul-2000  
C:Accession: I50222  
R:Funahashi, J.; Sekido, R.; Murai, K.; Kamachi, Y.; Kondoh, H.  
Development 119, 433-446, 1993  
A:Title: Delta-crystallin enhancer binding protein delta ERF is a zinc finger homeodomain  
A:Reference number: I50222; MUID:94116444; PMID:7904558  
A:Accession: I50222  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-114 <FUN>  
A:Cross-references: GB:D14313; NID:g391631; PIDN:BAAC3259.1; PID:g391632

Query Match 3.6%; Score 248.5; DB 2; Length 114;  
Best Local Similarity 18.8%; Pred. No. 2.4e-06;  
Matches 210; Conservative 144; Mismatches 377; Indels 385; Gaps 52;

QY 378 NSNGIPALQSSDGLGW-----QDKITVKG-----DTPGVGVPIKPLDSSRQN 426  
Db 22 NYNVEIA--NSDSODEKLHIVEESITDAACDASVPEDDLPDHTV--LFENESREG 77  
QY 427 GTATSYWCKFCSPSCSSSSSLK--LEHYGKHGAQVQSGGLNPELNKLRGVSINONDL 486  
Db 78 S-N-----SC-----WEDEGKETKEI-----LGPE----- 97  
QY 487 AKSSEGETWTKDKSSGAKKDFSSKGAEDNMVTSYNCQPCDFRYKSHGPDVIVGPL 546  
Db 98 AQSDVEGCTVKEDECDSDAENQNDHPNVEFL-----QEDTAVYPE 141  
QY 547 LRHYQQLHNIHKCT-----INHCPCPRG---LCSPEKHGEITYPFACKS 590  
Db 142 APEEDORQGTPEASQDENGTPDAFSQLLTCPYCDRGYKRTSLKEH---KYRHEKNE 198  
QY 591 NCSHCALLLHSPGAGSSRVKHQCHQCSFTTPDVVLLPHYESVHESQADYK--QEA 648  
Db 199 NFS-----CSLCSYT-----FAYR-QLDRHMTSHKSGRDQ 228  
QY 649 NHLQSDGQGSVKESKEHSCTKCDFTITQVEEISRYHRAHS---CYKCRQCSFTAAQTQ 705  
Db 229 RHVTQSSGNRKFK-----CTEGKAFKYKHLKHL--RIHSGEKYPCPNCKKXFSHSG 281  
QY 706 SLLEHNTVHCQEQDITANGEDGHAISTIKBPKIDFRVNNLTP--DSKMGEPVSES 763  
Db 282 SYSSHISKKK-----IGLMPVKGRARSLKTSQCSPSLASPSPARPQ 327  
QY 764 VVKREKLEEKDGLKE-----KYWTE-----SSDDLNVTVRGAD--LRGS 803  
Db 328 I--RQIKENKP-LQEGLPVNIQKTEPVYEFKPIVVASINGNSTPQNGVFGSGPLQAT 384  
QY 804 PS---YTOA-----SLGLTPVS-----GTOEQTKT 826  
Db 385 SSPQGVQAVLPTVGLVSPIS--NLSDIONLVKVAVDGNVPCVLENNEANLASKEQ-ET 443  
QY 827 LRDSPNVEAAH-----LAPPIVGLAVETKGFQAGAPAGEKSGAJPQOYP 871  
Db 444 ISNASTQAGHSLISAISLPVDDQDTTKIINYSLEQPSQLQVVPQNLKHSYV----- 499  
QY 872 ASGENKSKESQSLRRRGSGVFCANCLTTKSLWRKNAGG-----YVNCACGLYCK 925  
Db 500 ---TNSCKMEK-----LPEDLTIVSEK--DRNFEGETNDS--CLCDDC----- 537  
QY 926 LHSTPRPLNIKKONGEQIIRRTTRKRLNPEALQABQLNKQQRGSNEOVNPSLERRSE 985

Db 538 ----POCLNALQE-----LKHVETK--NPPQL-----POSSGTEAREKPSAPSETGE 579  
QY 986 DHUTESHOREIPLPS-LSKYBAOGLTKSHSAOCPVLVSQT-----LQHK-----RMOPLH 1036  
Db 580 NNLSPG---QPP-LKVLJLLKAYALNAQPSAEELSKIADSVNLPJDVVKWFEKWAQGC 636  
QY 1037 LQIKS-----POES\*GDPCGNSSSVSEK-----GSSER 1064  
Db 637 ISVQSSGSPSPPEQVXISSPTONEDQAATTNESEPN\*NNSONPANTSKSQTSSGGSTQN 696  
QY 1065 GSPIEKYMPAKHP-----NYSPPGSPTEKVOY?--FGLPVHNDQFSEADWL 1111  
Db 697 GS---RSSTPSPFNLSSRSNQGTYTAEGVQBEQPMEPDLSLPKOPGELLERSTIT 753  
QY 1112 RWSKYKLSVPGNPHYL-----SHVFGLPNPCNVYVPYPTFNLP 1151  
Db 754 SVYQNSVSVQVEFPNLITCAKKEPOKENSITDSDFIVNVIFESANPINIAIPVTAQLPT 813  
QY 1152 HFSVQSGNDIP-----LQLA-KHSRPGPT-ANGASKK-- 1184  
Db 814 -VAIADQNSVPCLRALAANKOTILPCVAYTYSTTVFAVQETPPKQTQANGSQDERQD 872  
QY 1165 --TKAPPNNVKNESPLNVKTEKVDSTODELSTKCVHCGIVFLDEVVVALHMSCHDSQP 1242  
Db 873 TSSGVSINVEDNDSDS:PPKXKXKTKENGXVA-CDLCKIFQKSSSLRKHYEHTGKRP 931  
QY 1243 FCCS\*CGHLCTDKYDFTTHIQRLHNNNAQVEKNC 1278  
Db 932 HEGGICKKAFKXKHL--EHRM-LHSGEKPYQCDKCK 967

## RESULT 4

C41602  
transcription factor GATA-2 - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 10-Jul-1992 #sequence\_revision 10-Jul-1992 #text\_change 26-Aug-1999  
C:Accession: C41602  
R:Zot, L.F.; Mather, C.; Burgess, S.; Boice, M.E.; Harland, R.M.; Crkin, S.H.  
Proc. Natl. Acad. Sci. U.S.A. 88, 10642-10646, 1991  
A:Title: Expression of GATA-binding proteins during embryonic development in Xenopus 1  
A:Reference number: A41602; MUID:92073343; PMID:1961730  
A:Accession: C41602  
A:Molecule type: mRNA  
A:Residues: 1-452 <ZON>  
A:Cross-references: GB:M76564; NID:9214170; PIDN:AAA49723.1; PID:9214171  
C:Superfamily: transcription factor GATA-2; GATA-type zinc finger homology  
C:Keywords: DNA binding; nucleus; transcription regulation; zinc finger  
F:264-317/Domain: GATA-type zinc finger homology <GZF1>  
F:267-291/Region: zinc finger GATA motif  
F:318-371/Domain: GATA-type zinc finger homology <GZF2>  
F:321-345/Region: zinc finger GATA motif

Query Match 3.5%; Score 241; DB 2; Length 452;  
Best Local Similarity 27.8%; Pred. No. 1.8e-06;  
Matches 99; Conservative 39; Mismatches 100; Indels 118; Gaps 19;

QY 750 LTPDSKMGPEVSEVVKREKLEKQGLKEKWTESDOLRNVTWRGACILRGSPSYTCA 809  
Db 163 VSPDPGAPSPSS-----RLEKOSIKYQM--SLSEGMK--MEGSPPLRS----- 235  
QY 810 SLGLLTPVSGTCEQT-KTLRDSFN-VEAAHLARP-YGLAVETKGFQAGAPAGEKSGALP 867  
Db 206 ----LAPM-GTQCS:HPPTYPVYVPAH-----DYSSGLPHFGSLGCPA-----SSSTP 252  
QY 968 QQ-----YPSAGENKSKESQSLR 887  
Db 253 KQPSKSESSEGCVCNCGATATPLWRDGTGHYLCNACGLYKXNGQNRPLKEKRLS 112  
QY 988 RRGSGVFCANCLTTKSLWRKNAGGVVNCAGLYQKLHSTPRPLNITKONGSQIARR 947  
Db 313 AAPRAGTCCANCTSTTLWRNRNAGCPVNCAGG-VYKLHKNRPLPTKKKEG-----CT 367  
QY 946 RTKELNPEALQAEQLNKQQRGSN-EE-----QVNGSPLEERSSEDHUTESHQREIFLPS 1000

```

Db 368 RNRKMSN-----KSKKKKSGCEFEELSRRCQKSGFFSAAA-----LASHM--APNGH 414
Qy 1001 LSKYEAGSLTKSHSAQPVLSVQTLDIHKRMQPLHIQIKSPQESTGDPGNSSSVS 1056
Db 415 LAFPSHSG---HILQTP-----PIH---PSSSLFGHPHSHSVMT 449

RESULT 5
A56355
DNA-binding protein Ikaros form 1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change 01-Dec-2000
C:Accession: A56355
R:Moinar, A.; Georgopoulos, K.
Mol. Cell. Biol. 14, 8292-8303, 1994
A:Title: The Ikaros gene encodes a family of functionally diverse zinc finger DNA-binding proteins
A:Reference number: A56355; MUID:95059058; PMID:7969165
A:Accession: A56355
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-518 <MOL>
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: alternative splicing; DNA binding

Query Match 3.5%; Score 239.5; DB 2; Length 518;
Best Local Similarity 23.7%; Pred. No. 2.7e-06;
Matches 128; Conservative 66; Mismatches 198; Indels 149; Gaps 26;

Qy 814 LTPVSGTGTQTKLRD-SPNVEAHLARPIYGLAVETKFLQGAFA---GGKSKALFQQ 869
Db 37 LSTTSQAQNSKSDRGASNVK-----VETQSENGRACEMNGSECAEDLRM 84

Qy 870 YPASGENKS---KDESOLLRRRRGGV-----FCA-NCLTTKTSLRKKYAK 912
Db 85 LDASGEKNGSHRDQSSAL-----SOGVGRLPNGLKXCD-CGIVCIGPVLMYVHKSH 139

Qy 913 GG---YVNCAGL-----YQKLSHTRPLNIIKONNGEIRRRTRKLNPEALQ 959
Db 140 TGERPFECNOCGASFTQGNLRLHILKHSQKP---FKCHLCNYACRRP-----DALT 189

Qy 960 AEQLNKQORGSGNERQVNGSPLEERS--EDHLTESHQ--REIPLPSL-----SKYBAQ 1007
Db 190 GLHRTSHVGPHKCCYCRGYKQSRSLSEHKERCHNYLESMLGPMYPIKETHNEVA 249

Qy 1008 GSLTKSHSAQPVLSVQTLDIHKRMQPLHIQIKSPQESTGDPGNSSSVSEKGSSEKSP 1067
Db 250 EDLCKIGAERSLVLDRLASNVAKRKSSM-----PKFLGDKCLSDMPYDSANYEKEDMM 303

Qy 1068 IEKYMRPAKP-NY-----SPRGS-----PIEKYQY-----PLFGLPFVHNDQ 1105
Db 304 TSHVMDQAINNAINYLGAESLRPLVQTTPSGSEVVPVSSYQLHKPPDCGPPRSHSAQ 363

Qy 1106 SEADMLRFWSKYKLSVCPNPHYLSHVPLG-LP-NPCQNYVP-----YPTFNLP 1151
Db 364 DAVDNLLLSKAK-SVS-----SEREASPSNQDSDTESNAEQRSGLIYLTNHNP 416

Qy 1152 HFSAVGSDNDIPLDLATKHSRPGPTANGASKETKAPPNVKNEGP-NVYKTEKVDNRSTQD 1211
Db 417 H-----ARNGLAKBEQRAYEVLRAASENSODAFRVYSTGE 453

Qy 1212 ELST-KCVHCGIVFLDEVMYALHMSCHDSG---PFQCSICQHLCKDYDFTTHIQGLH 1267
Db 454 QLVYKCEHCRVFLDHWMTYIHMGCHGCHGRDPFCNMGVHSODRVEFSSHITRGEH 513

Qy 1268 R 1268
Db 514 R 514

RESULT 6
S24047
zinc finger protein zfy-1 - mouse

```

```

K:Alternate names: Probable testis-determining factor
C:Species: Mus musculus (house mouse)
C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 01-Dec-2000
C:Accession: S04047
R:Ashworth, A.; Swift, S.; Afara, N.
Nucleic Acids Res. 17, 2864, 1989
A:Title: Sequence of cDNA for murine zfy-1, a candidate for Tdy.
A:Reference number: S04047; MUID:89240049; PMID:2497440
A:Accession: S04047
A:Molecule type: mRNA
A:Residues: 1-782 <ASH>
A:Cross-references: EMBL:X14382; NID:955478; PIDN:CAA32552.1; PID:955479
C:Genetics:
A:Gene: zfy-1
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: DNA binding; transcription regulation; zinc finger

Query Match 3.4%; Score 234; DB 2; Length 782;
Best Local Similarity 18.8%; Pred. No. 1e-05;
Matches 146; Conservative 93; Mismatches 258; Indels 278; Gaps 34;

Qy 20 LLEPIGTESKYSKVKESFSAQMSSENTQSDAAELNKKHEHSLHVQPPSSSKKD----- 74
Db 179 VLDSOMPLEQCDNDRNCEDYLXMSLDEPSKTLGSSSEVTNKAESSTGSKLCEASPE 238

Qy 75 LKSAVLSEKAGFN-----YESPSKGGNFPFPHDEVTRNMLAFSPFAAGVCEPL 125
Db 239 VIKVCILKADSEVDVGTIHAVERSETKNGK-----EAEVTDQS-TGIRVPRVNVKYSAS 292

Qy 126 KSPQRAE-----ADCFQ-----DVACTSPSGSLGFK-EDQ 154
Db 293 DSKCKSEEDTEVTVGSDSDAGGTAAUTPEHEQOMVSEIKAAFLPAWTAAYENNSDELEDQ 352

Qy 155 KMSPKATEETGCAQGGQANCCQLSPV-----SVASKNQVFSDDGVRLNKSKTDLVNDNP 210
Db 353 NVTASAL-LKQGES-----GLDAPVKCKSKKKRPS-----SKCYKSAIFV----- 393

Qy 211 DPAPISPELDQFKCNQGYGYGNDPTDLIKHFKYHLGLH-----NRTQDAELD--S 262
Db 394 --APEQGTJRVVPCVFCGKKF-----KTKRFLKRTKHPKPEYLANKKYKHCTECYSTN 444

Qy 263 KILALHNVQFESHKDFOKANRSVFSGLQDINSRPLVLLNGTYDVGVTSGETFIGR 322
Db 445 KKLILHNVQFESHKDFOKANRSVFSGLQDINSRPLVLLNGTYDVGVTSGETFIGR 322

Qy 323 TPCQGNKTYFRCKFNFTYMGNSSTELSQHPLOTHPNKIKAKSLPSSEVAKPSEKNKS 382
Db 491 ----KGVNKTCKKFCFCD--YETAEQTLNHLHLVYH-----RKK 523

Qy 383 IPALQSSDSDGLGK-----WCKKITVKAGDDIPVGVSVPIKP-LDSSRONGTEATSYW 435
Db 524 FPHI-----CGEGKGFRRPSALKKHIRVHTGE-----KP-----YE 555

Qy 436 CRFCSPSCSSSSSLKLEHYGKHQGAQVSGGLNPDLNKLRSQSVINONDLAKSSEGETM 495
Db 556 CQYCEYKSAQSSNLK--THIKSKH-----TSYNCQCFDFEYKSHGPDVYVVGPLLRH 549

Qy 496 TKTKDSSSGAKKKDS--SKGAEDNMV-----TSYNCQCFDFEYKSHGPDVYVVGPLLRH 549
Db 578 SKELPKKGCILLTTSOTKEAQOHAVLHOERSTHOCCHNKKSSNS-----SCKAKRH 629

Qy 550 YQGLNHNHKTIKHPFPFPRGLCSP---EKHLGEITYPFACRKNCSHCALLHLSPGA 606
Db 630 IISVHT--KAYPHKCDMCSKGFRPSBELKQHV-----A 660

Qy 637 AGSSRVKHQCHQCSFTTPDVLVLLPHYESVHSQASQVQKQEAHLQSDGQGSQSVKESKEH 666
Db 662 THKSKXQHQCHQCHDFNSDPFLSHHLSAH-----TKYVPF 697

Qy 667 SCTKCDFITQVEEISRHVRRAS---CYKRCQCSFTAAADTCSLLEHFNTHVCE 718
Db 698 KCKRCKKEFCQCELCQTH--YKTHSSRKVYQCEYCEYSTKQASGFKRHVTS:HTKD 751

```

```

RESULT 7
SC0647
finger protein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 31-Dec-1993
C:Accession: SC0647
R:Ruiz i Altaba, A.; Perry-O'Keefe, H.; Melton, D.A.
EMBO J. 6, 3065-3070, 1987
A:Title: Xfin: an embryonic gene encoding a multifingered protein in Xenopus.
A:Reference number: SC0647; PMID:88082679; PMID:2826129
A:Accession: SC0647
A:Molecule type: mRNA
A:Residues: 1-1350 <RUI>
A:Cross-references: EMBL:XC6021
A:Note: it is uncertain whether Met-1, Met-11, Met-16, Met-38, or Met-39 is the initiation
C:Genetics:
A:Gene: fin
C:Keywords: DNA binding; zinc finger

Query Match          3.4%; Score 233; DB 2; Length 1350;
Best Local Similarity 19.3%; Pred. No. 2.5e-05;
Matches 239; Conservative 138; Mismatches 39; Indels 472; Gaps 70;

QY 222 FKCNICGYYGNDPDLIKHPRKYLHGHNRTRDAELDSKILALHAKWQF----- 273
DB 354 YLCSHCNKGFGN--SDLVGHFRT--HTG--ERPQCAECHKGFIQKSLDLVHJLHRTHTGK 408
QY 274 ---SHSKDFQKVRNSVFGVLQDINSR2VLLNGTYDVQVTSGGTF-GIGRKTPDCC-- 326
DB 409 PFKSHC-DKKFTERSALAKHQRHTGKPY-----KSCDCKE 446
QY 327 -----QGNTRYCKFCNFYTMGNSSTLEQHLQHPNKKIKASLPSSEVAK 373
DB 447 FTQRNLIHQHRTGTERPYKCTLCDRTFIQNS--DLVKH-----QKVHANLPLSD-- 495
QY 374 PSEKNS-----NKSIPALQS-----SDSGDLGKQDKITV 403
DB 496 PHTANSPHKCKSDLTFSHWSTFMKSLHSGEKKFQCAEKKGFTQKSDLVK--HHRV 552
QY 404 KAGDTPGVGYPIKPL---DSRONGTEATSYW-----CKFCFSCESSSLK--LL 452
DB 553 HTGEK-----PFKCLLCKSPSQNSD--LHKHRIHTGKFPFCYCDXSTERSALI 603
QY 453 EHYGKHQGVQSGLNPELNDKLSRGSVINQNDLAKSBEGETMTKTDKSSGAKKDFSS 512
DB 604 KHRHRTHTG-----ERPHKCSVCQKGFQKSA-----LTKHSTRHTGK----- 641
QY 513 KQAEONNVTSYNCFDFRYSKSHGPDVIVVQPLRHVQOLNHIK---CT----- 560
DB 642 -----PYPTQC-----GKSPFQNSDLVKH-ORHITGKPYHCTECNKRFTG 693
QY 561 ---IKH-----CPFCPRGLCSPE---KHL-----GE-----ITYPFACRKS-- 591
DB 684 SSVLVKRRHSHGKPYRCQCEKTFIQSSDLVYH-LVWNGENPPAPATAFHEILIERNLT 743
QY 592 -----CSHCA-----LILLHLSPGAAGSRVKHCHQCS-----FTTPEDVDVLFPHY 633
DB 744 RSEPDYPCTEGKVFHQRPALLKHL---RTHKTEKRYPCNECDKSFQTSQD---LVKH 796
QY 634 ESVHESQASDVQKQAN--HLQSDGQOSVKESKEH-----SCTKCD--FTQVEEELSR 683
DB 797 LRHTHTGERPYHCECNKGFQNSD---LVKHQHTHTGERPYTCSQCDKGFQIR--SALT 851
QY 684 HYR---RAHSCYKRCOSFTTAADTQSLLEHFN-VHCOEQDITTANGEEEDGHA-STIKERPK 741
DB 852 HNRHTHTGKPYKCEQCKCFIQNSDLVYH-QRIH-----TGKPYHC----- 892
QY 742 IDFRVYNLLTPDSKMGEPVSESVVREKLEEK-----DGLKEKVTWTESSDDLNRVWFG 796
DB 893 -----PDCDKRFTGSSLLKHQRTHSRKIPYECGVCGKSPSQSS----- 931
QY 797 ADILRGSPSYTONSLGLTLPVSGTQE-----QTKTLRDSNVVEAAHLARPIYGLAVETKG 851

```

```

DB 932 -NLLKHLKCHSEQN---PPVALSSSLGFAVATQTHPDPVQ-----HI-----VYG---DTAS 976
QY 652 FLQCAPAGGKSGALPQQYPASGENSKDESQSLPRRRGSGVFCANCLTTKTSLWKNXA 911
DB 977 YI-SPEAAGERS----- 987
QY 912 NGGYVCNACS-----LYOKLHSTPRPLNIKQNGEQ:IRRTKRKLNPEALQAE 961
DB 988 ---FKCNDGCKCFAPRSVLIKHVRHTGERPKYC-----SQCTRSFIQKS 1029
QY 962 CLNKQKQSGNEECVNGSPLEERS--EDHLTESHQREIPLPSLSKYEAQCSLT:KSHSAQGP 1019
DB 1030 DLYKHRTHTGERPKYKGLCERSFVSKSALSQR-----VHKVEEP 1071
QY 1020 VLYSQDLDIHKMCPJH:CIKSFQESTGDFGNSSSVSEKGSSEKSGSP--IEKYVRPAKH 1077
DB 1072 VL-----NSAME--CCQVYWGSEKCDP--NSLVPQLEVIKEBESPH:VNAYSPLSL 1120
QY 1078 PNYSPF-----GSPTEKYQYVFLGLPFFVHNDQFSEADWLRFMSKYKLSVPGNPHYLSHVP 1132
DB 1121 QSYFPPILPEKGTG--RYSCEGCKCFTHKSV-----FJKNWR-----MHTG 1160
QY 1133 GLNPFQ-----NVYPYTFNL2PHFSVAGSDNDIP:DLAIKHSRPGGTANG 1179
DB 1161 EQPYTCCKEKGSPSQSALVKHVRHTGKPYACSTCGKSF:QKSDLA-KHOR-----HT 1215
QY 1180 ASKEKTKAPFNKVEGPNVVKTEKVDRTSTQBELSTKCVHCGIVF:DEVVYALHMSCHGD 1239
DB 1216 GERPYCTVCGKFFIDRSSVVKH---SR:HTGERPKYKNECT:KGFVCKSDLVKHMRTHTG 1272
QY 1240 SGFFQSCICHLCTDYDFTTHICRGLHNRNNAOVENKNGKP 1279
DB 1273 EKPYGNCDDR-----SESTHSASVRHQ---RMCNTGRP 1303

RESULT 8
A36389
transcription factor NF-E1b - chicken
N:Alternate names: DNA-binding protein: NF-E1b
C:Species: Gallus gallus (chicken)
C:Date: 28-Mar-1991 #sequence_revision 28-Mar-1991 #text_change 26-Aug-1999
C:Accession: A36389; S15645
R:Yamamoto, M.; Ko, L.C.; Leonard, M.W.; Beug, H.; Orkin, S.H.; Engel, J.D.
Genes Dev. 4, 1650-1662, 1990
A:Title: Activity and tissue-specific expression of the transcription factor NF-E1 mul
A:Reference number: A36389; PMID:91065513; PMID:2249770
A:Accession: A36389
A:Molecule type: mRNA
A:Residues: 1-466 <YAM>
A:Cross-references: EMBL:X56930; NID:G62965; PIDN:CAA40252.1; PID:G62966
C:Superfamily: transcription factor GATA-2; GATA-type zinc finger homology
C:Keywords: DNA binding; nucleus; transcription regulation; zinc finger
F:279-311/Domain: GATA-type zinc finger homology <GZFI>
F:261-305/Region: zinc finger GATA motif
F:332-385/Domain: GATA-type zinc finger homology <GZP2>
F:335-359/Region: zinc finger GATA motif

Query Match          3.2%; Score 222.5; DB 2; Length 466;
Best Local Similarity 28.1%; Pred. No. 2.3e-05;
Matches 93; Conservative 37; Mismatches 138; Indels 63; Gaps 14;

QY 750 LTPDSKMGEPVSSVVKREKLEKGLK-----EKWTESSDDLNRVWTKGAD----- 798
DB 172 VSPDPNSTAASPSSAGARQEDKDSIKYCVS:SEGKMKMESAPLSRLTSMGACQESTHS 231
QY 799 ILGSSSYTOA-----SLGLTTPVS--GTQETKTLRDSNVVEAAHLARPIYGLAVETKGF 852
DB 232 PIPTYFSYVFAAHDYSSJFHPQGFJGGPASSFTPRSKARSCSREGCVKNGCATATPL 291
QY 853 LOGAPAGGKSGALPQQYPASGENSKDESQSLRRRRSGVFCANCLTTKTSLWKNAN 912
DB 292 WRDGTGTHYLCNACGLYHKKNQGNRELKPKRPLSAARRAGTCCACCCCTTTTLWRNAN 951

```

```

Db 474 SSGVSAJDQKAISEMPSEVPEGGI---NLA$ALGVV$NAIKCTP-----SPKQ$N 524
QY 579 G$ITYPPACR$KNSCH$ALLHL$P$GAAG$SRVK-HQCHQCSFTT?DVVZLPHYESVH 637
Db 525 GE-----CR$SS-----SGKIKFKCKQCG-----H 545
QY 638 ESQASDVVQ$ENH:Q$GSDGQ$SVKESKEH$CTKCDITQVSE$SHYRR--AH$CYKCR 695
Db 546 Q$-SK$D-----QW$ARH$T$PA$KQLACQ$CNFVTEYKHLEVYRHNHIG$KPPCK 598
QY 696 QCSFTAACQ$LL$EHP--NTHVCEQDIT?ANG$EDGH$AISTIK$E$K?C?DFW$NLL?PD 753
Db 599 K$RYNCY$N$M$N$M$K$H$T$N$Y$FR$M$D$CY$ATK$YCH$U-----KLEL$KYN----- 646
QY 754 SK?G$E$V$E$SVYK$R$K$E$K$D$GL$K$E$K$Y$T$E$SS$D$D$LR$N$V$T$W$R$G$AD$IL$R$G$S$Y$T$Q$AS$L$G$L 813
Db 647 ---H$R$V$E$G$I-----E$M$G$D$-----S$P$P$T$S$D$AT$I 672
QY 814 ?V$G$T$G$E$Q$KTL$D$S$N$V$E$A$H$L$R$P$I---V$L$A$V$E$T$G$F$L$C$G$A$P$A$G$E$K$S$G$AL$P$Q$Y 870
Db 673 F$P$L$K$Q$B$K$I$T$E$V$E$P$T$S$T$A$Q$P$P$F$M$X$M$G$L$N$F$AN$H$M$L----- 715
QY 871 P$S$G$N$K$D$E$S$Q$S$L$R$R$R$G$S$G$V$F$C$A$N$C$U$T$T$K$L$W$R$K$N$G-$G$Y$V$N$A$C$G$L$Y$Q$M$L$H$T 929
Db 716 ----NKHLDVGLMGLR$N$V$S$PLKCSACD$F$V$A$S$A$D$E$K$W$H$S$M$H$L$N$S$N$V$T$S$I$A$S$ 770
QY 930 P$P$L$N$I$K$Q$N$G$E$Q$?R$R$T$R$K$N$P$E$A$Q$E$Q$L$N$K$O$O$R$G$N$E$Q$V$G$N$G$P$L$E-----R$R 983
Db 771 Y$N$S$L$M$-----P$P$S$H$V$A$P$D-----N$N$A$L$E$S$Y$D$C$D$V$K$ 799
QY 984 S$C$H$L$T$E$S$H$Q$E$?P$L$P$S$?S$K$Y$E$A--Q$S$L$T$K$H$S$A$Q$P$V$L$V$S$T$L$C$H$K$M$Q$P$H$?C$I$K$S 104
Db 900 D$U$N$T$E$S$H-----C$Y$E$X$D$Q$G$----- 817
QY 1042 P$O$E$T$G$D$C$N$S$S$V$S$E$G$S$S$E$R$G$S$P$E$K$Y$M$R$P$A$K-H$P$N$Y$P$P$S$P$E$K$Y$Q$P$F$L$G$L$P$E$V 1100
Db 818 ---D$A$V$S$P$T$G$S$Q$-S$S$G$E$E$T$K$K$S$L$S$L$E$C$I$S$A$R$A$N$G$N$S$P-----Y 858
QY 1101 H$N$D$F$O$E$A$D$W$L$R$F$W$S$K$Y$K$L$V$P$G$N$H$Y$S$H$V$P$G$P$N$P$C$Q$N$Y$P$Y$P$T$N$L$P$H$F$S$A$V----- 115
Db 859 S$N$D$A$V$E$K$D-----G$E$S$A$D$A$P$H$P$S$D$T$T$S$V$E$P-----P$L$H$S$S$I$V$A$S$I 999
QY 1157 ---G$S$D$N$I$P$L$D$A$I$K$H$R$G$P$-T$A$N$G$A$K$E$T$K$A$P$N$V$K$N$E$G$L$N$V$K$T$E$Y$D$R$S$T$Q$D$E 121
Db 899 P$T$P$Q$E$N$F$L$O$S$I$A$Q$S$L$L$P$L$A$N$R$P$----- 927
QY 1213 L$T$K$V$H$Q$G$?V$F$O$E$W$Y$A$L$H$V$S$H$C$P$S$G$P$P$Q$S?C$Q$H$L$C$T$D$K$Y$F$T$H?Q$R$G$L$H$R 1266
Db 928 -A$P$Y$C$H$R$K?P$D$T$C$V$L$D$H$M$R$F$H$T$P$G$N$P$E$N$C$S$C$O$Y$O$A$F$N$E$L$S$F$A$L$H$W$Y$Q$A$R$H$O 962

```

```

RESULT 10
S05548
gap protein: hunchback - fruit fly (Drosophila virilis)
N:Alternate names: finger protein, hunchback
C:Species: Drosophila virilis
C>Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 24-Sep-1998
C:Accession: S05548
R:Teier M., Pfeife, C., Tautz, D.
EMBO J. 9, 1517-1525, 1989
A>Title: Comparison of the gap segmentation gene hunchback between Drosophila
A:Reference number: S05548; MJID:89356623; PMID:2504591
A:Accession: S05548
A:Molecule type: DNA
A:Residues: 1-816 <TR>
A:Cross-references: EMBL:X:5359; NID:g9178; PID:g9.79
C:Genetics:
A:Gene: hb
A:Cross-references: FlyBase:FBgn0013116
C:Keywords: DNA binding; zinc finger

Query Match 1.02% Score 218.5 DB 2: Leuc3h 816:

```





```
RESULT 12
157561
transcription factor GATA-4 - human
C:Species: Homo sapiens (man)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 17-Mar-1999
C:Accession: 157561 J04020
R:Yamagata, T.; Nishida, J.; Sakai, R.; Tanaka, T.; Honda, H.; Hirano, K.; Mano, H.; Yazaki, M.;
Mol. Cell. Biol. 15, 3830-3839, 1995
A:Title: Of the GATA-binding proteins, only GATA-4 selectively regulates the human inter-
A:Reference number: 157561; MUID:95311982; PMID:7791790
A:Accession: 157561
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-439 <RES>
A:Cross-references: GB:S78666; NID:G1042085; PID:G1042086
R:Huang, W.Y.; Cukerman, E.; Liew, C.C.
Gene 155, 219-223, 1995
A:Title: Identification of a GATA motif in the cardiac alpha-myosin heavy-chain-encoding
A:Reference number: JC4020; MUID:95237613; PMID:7721094
A:Accession: JC4020
A:Molecule type: mRNA
A:Residues: 'MAANHGPAGAYAGGPGFMHAGAGASS', 'PVYLPTRVFESSVGLSLVLOGGAGSAGG', 64-68, 1994
A:Cross-references: GB:L34357; NID:9508483
A:Experimental source: heart
C:Comment: This protein regulates a set of cardiac-specific genes and plays a crucial ro
C:Genetics:
A:Gene: GDB:GATA4
A:Cross-references: GDB:433278; OMIM:600576
A:Map position: 8p23.1-8p23.1
C:Superfamily: transcription factor GATA-4; GATA-type zinc finger homology
C:Keywords: DNA binding; transcription factor; zinc finger
F:211-264/Domain: GATA-type zinc finger homology <GZF1>
F:214-238/Region: zinc finger GATA motif
F:265-318/Domain: GATA-type zinc finger homology <GZF2>
F:268-292/Region: zinc finger GATA motif

Query Match 3.1%; Score 215; DB 2; Length 439;
Best Local Similarity 33.0%; Pred. No. 5.6e-05;
Matches 69; Conservative 22; Mismatches 68; Indels 50; Gaps 9;

QY 873 SGENKSDSOSLLRRRGSGVCANCLTTKTSIMRWKANGGVYCNACGLYCKLHSTPP 932
D 245 NGINRPLIKPQRLSARRVGLSCANQTTTLWRNAGEPVCNACGLYKLRGVPR 304
QY 933 LNIQKNGEQIIRRTTRKLNFEALQAEQINLQQR-----GSNEEGVNGSPLRRSEDL 968
D 305 LAMRKEG-----IQTRKKPKN-----LNKSKTPAAPSGESLPPASGASSNSNAT 351
QY 989 TESHQREIPL---PSLSKYEAQSLTKSHSAQPVLSQTL-----DIHKMQPL 1035
D 352 TSSSEERPKTFGLSSH-----YGHSS-----VSQTFVSAMSGHGPSHPVLSAL 400
QY 1036 HIQIKSPQSTGPGNSSSVSEKGSER 1064
D 401 KL---SPQ-----GYASPVQSPTSSK 420

RESULT 13
A40815
transcription factor GATA-2 (version 1) - human
C:Species: Homo sapiens (man)
C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 26-Aug-1999
C:Accession: A40815
R:Lee, M.E.; Temizer, D.H.; Clifford, J.A.; Quertermous, T.
J. Biol. Chem. 266, 16188-16192, 1991
A:Title: Cloning of the GATA-binding protein that regulates endothelin-1 gene express-on
A:Reference number: A40815; MUID:91340773; PMID:1714909
A:Accession: A40815
A:Molecule type: mRNA
A:Residues: 1-480 <LEE>
A:Cross-references: GB:W6889; NID:G192995; PID:AAA35868.1; PID:G182996
C:Genetics:
```

```
A:Gene: GDB:GATA2
A:Cross-references: GDB:133757; OMIM:137295
A:Map position: Xp11.23-Xp11.23
C:Superfamily: transcription factor GATA-2; GATA-type zinc finger homology
C:Keywords: DNA binding; nucleus; transcription regulation; zinc finger
F:292-345/Domain: GATA-type zinc finger homology <GZF1>
F:293-349/Region: zinc finger GATA motif
F:346-399/Domain: GATA-type zinc finger homology <GZF2>
F:349-373/Region: zinc finger GATA motif

Query Match 3.1%; Score 215; DB 2; Length 480;
Best Local Similarity 28.0%; Pred. No. 6.4e-05;
Matches 96; Conservative 26; Mismatches 89; Indels 106; Gaps 14;

QY 771 EEKDLKEKV-WTESSDDLNNVTWRGADILRSPSYTQASLGLTPVSGTQCTKTLRD 829
D 206 EDKGVKVCVSTESMKVE-----SGSP-----LRPGLAT--YGTQ----- 239
QY 830 SPNVEAAHLARPIYGLAVETKG-----FLOGAPAGEKSGALPQ----- 869
D 240 ----PATHHPITPYESVYVPAAAHDYSSGJFHPGGJLGGPASSFTPKORSKARSCSEGREC 295
QY 870 -----YPASGENKSKDESQSLRRRRRRGGSGVFCANCLTT 902
D 296 VNCGATATP-WERDGTGTY-CNACGLYKMGNGQNRPLIKRPLSAARRAGTCCANCTT 355
QY 903 KTSLWRKXNANGVYCNACGLYQKLSHTPRPLNLIKQNGEQIIRRTTRKLNFEALQAEQ 962
D 356 --TLKRRNANGDPVCNACGLYVYKLNVRPLTKKEG-----QTRNRKXSN-----K 403
QY 963 LNKQQRSGN--EE-----QVNGSPLRRS-----EDHJTE-SHQREIPLPSLSKYEAQ 1007
D 404 SKSKGAGCTFEELSKCKQKSEFFSAAA-AGHMAPVGHILPFPFSSHSHI-LPTPTPIHPS 462
QY 1008 GSLTKSH 1014
D 453 SSLSFGH 469

RESULT 14
A56229
lymphoid transcription factor Ikaros/IyF-1, form IV - mouse
N:Alternate names: Ikaros/IyF-1 form I; Ikaros/IyF-1 form II; Ikaros/IyF-1 form III
C:Species: Mus musculus (house mouse)
C>Date: 03-Nov-1995 #sequence_revision 03-Nov-1995 #text_change 01-Dec-2000
C:Accession: A56229
R:Haem, K.; Ernst, P.; Lo, K.; Kim, G.S.; Turck, C.; Smale, S.T.
Mol. Cell. Biol. 14, 7111-7123, 1994
A:Title: The lymphoid transcription factor IyF-1 is encoded by specific, alternative-y
A:Reference number: A56229; MUID:95021239; PMID:7935426
A:Accession: A56229
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-392 <HAH>
A:Cross-references: GB:S74517
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: alternative splicing

Query Match 3.1%; Score 213; DB 2; Length 392;
Best Local Similarity 21.4%; Pred. No. 6.3e-05;
Matches 122; Conservative 65; Mismatches 169; Indels 214; Gaps 26;

QY 728 EDGHAISTI--KEEPKIDFRVYNLLTPDSKMGEPVSVVYKREKLEKGLKEKWTSS 785
D 4 DEGDMSQVSKESPPVSC-----TPDEG-----DEPMVPVEDLSTTSGAQQ----NSK 48
QY 786 SDDLNNVTWRGAD-----IL--RGSPSYTQASLGLLTPVSGTCEOTK-LRDSPNVEAA 836
D 49 SD--RGVAVYAGDGRDFHAIISDRGMS-----NVKVCSDENGRAC 91
QY 837 FLARPIYGLAVETKGF-QGAPAGKEKSGALPQQVYPASGENKS----KQESQSLRRRRGSG 893
D 92 EM-----NGEECAEDLRMLDASGEKMGSHRQGGSAJ-----SG 126
```

```
QY 894 VFCANLTTKTSLWRKNANGVYVCNACGLYQKHLSTPRPLN::KQNGEQIIRRRTRKEL 953
Db 127 VGGI-----R.PNGKLTCDICGIV-----CI 147
QY 954 NPEALQAEQLNKQORGSNEOVNGSP:ERRSEDHLTESHORE:PLPSLSKYEACGSLTKS 1013
Db 148 GPNVLM-----VHKRSH---TKKCLSDMPYS-ANYEKEDNMT-S 182
QY 1014 HSAQQPVLSQTLDIHKRMOP:HIQIKSPOESTGDFGNSSSVSEGGKSSERGSP:EKYKR 1073
Db 183 HVMQAINNAINVYLGAEPL---VQTP-----PGSEVVVPVSSMYC 223
QY 1074 PAKHPNYPGSP:IEKYQYPLGLPVFHNDQSEADWLRFWSKYKLSVPCNPHYLSHVPF 1133
Db 224 LHKPPSDGPPRS-----MHSQAQDAVENLLLSKAK-SVS-----SREFA 261
QY 1134 LP-NPCQNVVP-----YPTFNLPHPFSVAGSDNDIP:LDIAIKHSRPGTKAG 1179
Db 262 SPNSQCDSTDTESNAEQRSGLIYLTNHPH-----ARNG 298
QY 1180 ASKEKTKAPENVKNEGLNVVKTEDRSTQDELST-KCVHCGIVFDEVMYALHVSCHG 1238
Db 299 LALKEQRAVEVLRAASENSQDAFRVSVTSGEQLKVYKCEHCRVLF:LDHVMYTIHMGCHG 358
QY 1239 DSGPPQCSICQHLCTDKYDFTTHIQRLHR 1268
Db 359 FRDPFECMCGYHSQDRYEFSSHITRGHR 388

RESULT 15
S40382
box A-binding factor - fruit fly (Drosophila melanogaster)
N;Alternate names: ABF; transcription factor 3GATAB
C;Species: Drosophila melanogaster
C;Date: 13-Jan-1995 #sequence_revision 06-Sep-1996 #text_change 16-Jul-1999
C;Accession: S40382
R;Abdel, T.; Michelson, A.M.; Maniatis, T.
Development 119, 623-633, 1993
A;Title: A Drosophila GATA family member that binds to Adh regulatory sequences is expressed in the developing eye
A;Reference number: S40382; MUID:94244465; PMID:8187633
A;Accession: S40382
A;Molecule type: mRNA
A;Residues: 1-779 <ABE>
A;Cross-references: EMBL:X76217; NID:9441491; PIDN:CAA53607.1; P/D:9441492
C;Comment: This transcriptional activator is the earliest known marker of the developing eye
C;Genetics:
A;Gene: FlyBase:srp
A;Cross-references: FlyBase:FBgn0003507
C;Superfamily: box A-binding factor; GATA-type zinc finger homology
C;Keywords: DNA binding; nucleus; transcription factor; transcription regulation; zinc finger
F;316-369/Domain: GATA-type zinc finger homology <GZF>
F;319-343/Region: zinc finger GATA motif

Query Match 3.1%; Score 213; DB 1; Length 779;
Best Local Similarity 25.3%; Pred. No. 0.00016;
Matches 80; Conservative 38; Mismatches 98; Indels 104; Gaps 12;

QY 849 TKGFLQCAPAGGKSGALPOQYPASG-----ENKSKDESQSLRRRRRGSGVFCAN 898
Db 263 TSGAASSYSCPGSNATSAATSAVASGTAAATATLDEHVSRRRRLSASKR-AG:SCSN 321
QY 899 CLTTKTSLWRKNANGVYVCNACGLYQKHLSTPRPLN::KQNGEQIIRRRTRKELNFEAL 958
Db 322 CHTTHTSLWRNPAGEPVCNACGLYKLSVPRFLTKKDT-----IQKSKRK---PKGT 373
QY 959 QAEQ-LNKQQRGSNEEQVNGSFLER-RSEDHLTESHOREI-----P 997
Db 374 KSEKSKGSKNALNAIMEGSLVTNCHNVGVLDSSQMDVNDOMKPOLDLKPNYSYSSQP 433
QY 998 LPSLSKYEAQGSLSL-----TKSHSAQOPVLYS--- 1023
Db 434 QQLPQYQQQQQQLVNADQHSAAASSPHSMGTSLSPSAMSHQHTHPHQQQQQQLCSGWT 493
```

```
QY 1024 ----QTLQ-HKR-----MOPLHIQIKSP--QESTGDFGNSSSVSECKG 1060
Db 494 CRPTQTTXCRRRSTCCSS:SSSRRAACSTHFAHPJL:QHPSPHQLHNNNNNNSSLLFNNN 553
QY 1061 SSERGSP-----IEKYM 1073
Db 554 NNNSSNNNNNNK:JQKYLQ 573

Search completed: October 29, 2003, 11:44:08
Job time : 76 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2003, 11:29:56 ; Search time 39 Seconds  
(without alignments)

1544.646 Million cell updates/sec

Title: US-09-702-216-2

Perfect score: 6851

Sequence: 1 MTRKKQPLPLNVASBEGGI.....IORGJHRNNAQVEKNGKPK 1281

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6843	99.9	1281	1 TRPS_HUMAN	Q9uf77 homo sapien
2	6424	93.8	1281	1 TRPS_MOUSE	Q925h1 mus musculus
3	5107	74.5	1271	1 TRPS_XENLA	Q902s6 xenopus lae
4	273.5	4.0	757	1 HUNB_DROSE	O62538 drosophila
5	271	4.0	758	1 HUNB_DROSE	P05084 drosophila
6	269	3.9	519	1 IKAR_HUMAN	Q13422 homo sapien
7	266.5	3.9	767	1 HUNB_DROOR	O62537 drosophila
8	251.5	3.7	1829	1 Y296_HUMAN	O15015 homo sapien
9	248.5	3.6	759	1 HUNB_DROYA	O62541 drosophila
10	248.5	3.6	1114	1 TCF8_CHICK	P26197 gallus gall
11	245	3.6	3703	1 ABFI_HUMAN	Q15911 homo sapien
12	241	3.5	452	1 GAT2_XENLA	P23770 xenopus lae
13	238	3.5	1350	1 XFIN_XENLA	P08045 xenopus lae
14	234.5	3.4	971	1 AREA_GIBFU	P78688 gibberella
15	234	3.4	517	1 IKAR_MOUSE	Q03267 mus musculus
16	234	3.4	782	1 ZFY1_MOUSE	P10925 mus musculus
17	230	3.4	518	1 IKAR_CHICK	Q42410 gallus gall
18	223.5	3.3	1687	1 Z142_HUMAN	P52746 homo sapien
19	222.5	3.2	466	1 GAT2_CHICK	P23742 gallus gall
20	222.5	3.2	982	1 HBL1_CAPEL	Q9xyd3 caenorhabdi
21	221.5	3.2	817	1 HUNB_MOUSE	Q01178 musa domes
22	220.5	3.2	3726	1 ABFI_MOUSE	Q61329 mus musculus
23	218.5	3.2	816	1 HUNB_DROVI	P13361 drosophila
24	218	3.2	522	1 IKAR_ONCMY	O13089 oncorhynch
25	218	3.2	783	1 ZFY2_MOUSE	P20662 mus musculus
26	217	3.2	397	1 GAT5_HUMAN	Q9bwx5 homo sapien
27	215	3.1	442	1 GAT4_HUMAN	P43694 homo sapien
28	213	3.1	480	1 GAT2_HUMAN	P23769 homo sapien
29	213	3.1	779	1 SRP_DROME	P52172 drosophila
30	212.5	3.1	391	1 GA6B_XENLA	P70005 xenopus lae
31	212.5	3.1	440	1 GAT4_RAT	P46152 rattus norv
32	212.5	3.1	526	1 HELI_HUMAN	Q9uk57 homo sapien
33	212.5	3.1	526	1 HELI_MOUSE	P51183 mus musculus

#### RESULT 1

ID	TRPS_HUMAN	STANDARD:	PRT:	1281 AA.
AC	Q9UHF7; Q9NWE1; Q9UHH6;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Zinc finger transcription factor Trps. "Zinc finger protein. GC79;			
DE	Tricho-rhino-phalangeal syndrome type I protein).			
GN	TRPS1.			
OS	Homo sapiens (Human).			
CC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
CX	NCBI_TaxID=9606;			
RL	[1]_TaxID=9606;			
RN	SEQUENCE FROM N.A.			
RP	MEDLINE=2008282; PubMed=10615111;			
RX	Medline=2008282; PubMed=10615111;			
RA	Momeni P., Gloeckner G., Schmidt O., von Holtum D., Albrecht E.,			
RA	Gillieson-Kaesbach G., Hennekam R.C.M., Meinecke P., Zabel B.,			
RA	Rosenthal A., Hortschmcke B., Lueddecke H.-J.;			
RT	"Mutations in a new gene, encoding a zinc-finger protein, cause			
RT	tricho-rhino-phalangeal syndrome type I";			
RL	Nat. Genet. 24:71-74(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Prostate;			
RX	MEDLINE=20432376; PubMed=10974377;			
RA	Chang G.T.C., Steenbeek M., Schippers E., Blok L.C., van Weerden W.M.,			
RA	van Aalwijk D.C.G., Eussen B.H.C., van Steenbrugge G.J.,			
RA	Brinkmann A.O.;			
RT	"Characterization of a zinc-finger protein and its association with			
RT	apoptosis in prostate cancer cells";			
RL	J. Natl. Cancer Inst. 92:144-142(2000).			
RN	[3]			
RP	SEQUENCE OF 495-1279 FROM N.A.			
RC	TISSUE=Embryo;			
RA	Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,			
RA	Nasikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,			
RA	Tanase T., Nemura Y., Togiya S., Komai F., Hara R., Takeuchi K.,			
RA	Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,			
RA	Wakamatsu A., Makamura Y., Nagahara K., Masuho Y., Oshima A.;			
RT	"NEO human cDNA sequencing project";			
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	VARIANTS TRPS3 ASP-934; PRO-951; GLN-908; PRO-908 AND THR-919.			
RX	MEDLINE=21063177; PubMed=1112658;			
RA	Lueddecke H.-J., Schaper J., Meinecke P., Momeni P., Gross S.,			
RA	von Holtum D., Hirsch H., Abramowicz M.J., Albrecht B., Apacik C.,			
RA	Christen H.-J., Clauser U., Devriendt K., Pastnacht E., Forderer A.,			
RA	Friedrich U., Godtschap T.H.C., Greive X., Hamm H., Hennekam R.C.M.,			
RA	Rinkel G.K., Hoeftzenbein K., Kayserli H., Matjewska P., Mathieu M.,			
RA	McLeod R., Miro A.T., Moog J., Nagai T., Nikiawa K., Oerstavik K.H.,			
RA	Plechni E., Saitz C., Schmidtke J., Tranbjaerg L., Tsukahara M.,			
RA	Plechni E., Zabel B., Gillessen-Kaesbach G., Horstheke B.;			
RT	"Genotypic and phenotypic spectrum in tricho-rhino-phalangeal syndrome			
RT	types I and III";			

34	212	3.1	404	1	GAT5_MOUSE	P97489 mus musculus
35	212	3.1	480	1	GAT2_MOUSE	O09100 mus musculus
36	213	3.1	1507	1	PRDF_HUMAN	P57071 homo sapien
37	209.5	3.1	391	1	GA5A_XENLA	Q91678 xenopus lae
38	209.5	3.1	392	1	GAT4_XENLA	Q91677 xenopus lae
39	208	3.0	436	1	GAT3_BRARE	Q91428 brachydanio
40	208	3.0	1342	1	Z135_HUMAN	Q9P422 homo sapien
41	207.5	3.0	444	1	GAT3_CHICK	P23825 gallus gall
42	207	3.0	443	1	GAT3_HUMAN	P23771 homo sapien
43	207	3.0	443	1	GAT3_MOUSE	P23772 mus musculus
44	207	3.0	1109	1	TCF8_RAT	Q62947 rattus norv
45	207	3.0	1124	1	TCF8_HUMAN	P37275 homo sapien

#### ALIGNMENTS

Am. J. Hum. Genet. 68:81-91(2001).

[5] VARIANT TRPS3 GLN-908.

MedJine=21665516; PubMed=1807863;

Kobayashi H., Hino M., Shimodaira M., Iwakura T., Ishihara T.,

RA Ikekubo K., Ogawa Y., Nakao K., Kurahachi H.,

RA "Missense mutation of TRPS1 in a family of tricho-rhino-phalangeal

RT syndrome type III.";

Am. J. Med. Genet. 107:26-29(2002).

CC -!- FUNCTION: Transcriptional repressor. May act to restrict

CC expression of GATA-regulated genes at selected sites and stages in

CC vertebrate development. Might be involved in prostate cancer

CC apoptosis.

CC -!- SUBUNIT: Binds specifically to GATA sequences (By similarity).

CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).

CC -!- TISSUE SPECIFICITY: Ubiquitously expressed in the adult. Found in

CC fetal brain, lung, kidney, liver, spleen and thymus. More highly

CC expressed in androgen-dependent than in androgen-independent

CC prostate cancer cells.

CC -!- DISEASE: Defects in TRPS1 are the cause of tricho-rhino-phalangeal

CC syndrome type I (TRPS1) [MIM:190350], an autosomal dominant

CC disorder characterized by craniofacial and skeletal abnormalities.

CC It is allelic with tricho-rhino-phalangeal type III. Typical

CC features include sparse scalp hair, a bulbous tip of the nose,

CC protruding ears, a long first phalanx and a thin upper vermillion

CC border. Skeletal defects include cone-shaped epiphyses at the

CC phalanges, hip malformations and short stature.

CC -!- DISEASE: Defects in TRPS2 are a cause of tricho-rhino-phalangeal

CC syndrome type II (TRPS2) [MIM:150230]. TRPS2 is a contiguous gene

CC syndrome due to deletions in chromosome 8q24.1 and resulting in

CC the loss of functional copies of TRPS1 and EXT1.

CC -!- DISEASE: Defects in TRPS3 are the cause of tricho-rhino-phalangeal

CC syndrome type III (TRPS3) [MIM:190351], an autosomal dominant

CC disorder characterized by craniofacial and skeletal abnormalities.

CC It is allelic with tricho-rhino-phalangeal type I. In TRPS3 a more

CC severe brachydactyly and growth retardation are observed.

CC -!- SIMILARITY: Contains 7 C2H2-type zinc fingers.

CC -!- SIMILARITY: Contains 1 GATA-type zinc finger.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

EMBL; AF183810; AAF23614.1; -

DR EMBL; AF178030; -; NOT ANNOTATED\_CDS.

DR EMBL; AF264784; AAG21134.1; -

DR EMBL; AK000948; BAA91441.1; -

DR HSSP; P17679; IGNF.

DR Genew; HGNC:12340; TRPS1.

DR MIM; 604386; -

DR MIM; 190350; -

DR MIM; 190351; -

DR MIM; 150230; -

DR InterPro; IPR007087; Znf\_C2H2.

DR InterPro; IPR000679; Znf\_GATA.

DR Pfam; PF00320; GATA.1.

DR Pfam; PF00096; zf-C2H2.4.

DR PRINTS; PR00619; GATAZNFINGER.

DR SMART; SM00355; Znf\_C2H2; 9.

DR SMART; SM00401; Znf\_GATA; 1.

DR PROSITE; PS00344; GATA\_2N\_FINGER\_1; 1.

DR PROSITE; PS00114; GATA\_2N\_FINGER\_2; 1.

DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 2.

DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 1.

CC Transcription regulation; Repressor; Zinc-finger; DNA-binding;

KW Nuclear protein; Metal-binding; Repeat; Disease mutation.

KW 2N\_FING 222 247 C2H2-TYPE 1 (ATYPICAL).

FT 2N\_FING 333 358 C2H2-TYPE 2 (ATYPICAL).

FT 2N\_FING 614 637 C2H2-TYPE 3 (ATYPICAL).

666	689		C2H2-TYPE 4.
692	715		C2H2-TYPE 5.
896	920		GATA-TYPE.
1215	1237		C2H2-TYPE 6.
1243	1267		C2H2-TYPE 7.
894	894		V->D (in TRPS3; in heterozygous status
			has a milder effect causing TRPS1).
			/Frid=VAR_012807.
901	901		T->P (in TRPS3; severe).
			/Frid=VAR_012808.
908	908		R->P (in TRPS3; severe).
			/Frid=VAR_012809.
908	908		R->Q (in TRPS3).
			/Frid=VAR_012810.
919	919		A->T (in TRPS3).
			/Frid=VAR_012811.
115	115		S->F (in REF. 1; AAF23614).
1279	1279		NGXP->KKKT (in REF. 3).
1281	1281		AA: 141520 MW; 2157804F55EB71CC CRCE4;
			Query Match 99.9%; Score 6843; DB 1; Length 1281;
			Best Loca. Similarity 99.9%; Pred. No. 0;
			Matches 1260; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY	1	MVRKNPFLRNVASEGEGQILEPIGTESKVSGNKEFSACQXSENTQSCAAELNHKEEH	60
DB	1	MVRKNPFLRNVASEGEGQILEPIGTESKVSGNKEFSACQXSENTQSCAAELNHKEEH	60
QY	61	SLHVQDSSSSKKDLKSAVLSEKAGFNYESPSKGNFSPSPHDEVTDENMLASFPAAGG	120
DB	61	SLHVQDSSSSKKDLKSAVLSEKAGFNYESPSKGNFSPSPHDEVTDENMLASFPAAGG	120
QY	121	VCEPLKSPQRAEADDDPODMACTPSGDSLETKEQKMSFKATEETGQAQSQOANQGLSPV	180
DB	121	VCEPLKSPQRAEADDDPODMACTPSGDSLETKEQKMSFKATEETGQAQSQOANQGLSPV	180
QY	181	SVASKNPQVPSDGGVRUNKSKTLLVNDNPDPAPLSPELQCFKNCIGYGYGNDPTDLI	240
DB	181	SVASKNPQVPSDGGVRUNKSKTLLVNDNPDPAPLSPELQCFKNCIGYGYGNDPTDLI	240
QY	241	KKPKYHLGLHNRTRQDAELDSKTLALHNVCVSHSKCFCKVARSFVGLQGINSRPV	300
DB	241	KKPKYHLGLHNRTRQDAELDSKTLALHNVCVSHSKCFCKVARSFVGLQGINSRPV	300
QY	301	LLNGTYDVQVTSSTGTFIGIGRKTDCGCKTKYFRCKFCNFTYMGNSSTELSQHFLQHPN	360
DB	301	LLNGTYDVQVTSSTGTFIGIGRKTDCGCKTKYFRCKFCNFTYMGNSSTELSQHFLQHPN	360
QY	361	KIKASLPSSEVAKPSEKSNKSIIPALQSSDGGDLGKWQDKITVKGDCPTVGYSVFKPL	420
DB	361	KIKASLPSSEVAKPSEKSNKSIIPALQSSDGGDLGKWQDKITVKGDCPTVGYSVFKPL	420
QY	421	DSSRCNGTEATSYVWKPCFSCESSSSSLKLLPHYGHQHGAVQSGGNPELNDKLSRGSV	480
DB	421	DSSRCNGTEATSYVWKPCFSCESSSSSLKLLPHYGHQHGAVQSGGNPELNDKLSRGSV	480
QY	481	INQNDLAKSSEGETMTKTDKSSGAKKXFFSSKGAEDMMVTSYNCQCFDPRYSKSHGPDV	540
DB	481	INQNDLAKSSEGETMTKTDKSSGAKKXFFSSKGAEDMMVTSYNCQCFDPRYSKSHGPDV	540
QY	541	IVVGPILLRHVQQLHNIHKTIKPCFCPRGLCSPEKHLGEITYPFACRKNCSHCALLLL	600
DB	541	IVVGPILLRHVQQLHNIHKTIKPCFCPRGLCSPEKHLGEITYPFACRKNCSHCALLLL	600
QY	601	HLSPGAAGSRVHKHCQCSFTTPDVVLLFHVSVHESQASDVKQEAHLQSGSDGQSV	660
DB	601	HLSPGAAGSRVHKHCQCSFTTPDVVLLFHVSVHESQASDVKQEAHLQSGSDGQSV	660
QY	661	KESKEHSCTKCDFITQVDEEISRHRRASHCYKRCQCSFTAAQTQSLLEHFNTHVCOEQD	720
DB	661	KESKEHSCTKCDFITQVDEEISRHRRASHCYKRCQCSFTAAQTQSLLEHFNTHVCOEQD	720
QY	721	TTANGSEDGHA-STIKKEPKIDFRVYNLLTPUSKMGEPVSESYYKEXLEKQGLKEKV	780

```

Db 721 ITTANGEEDGHALSTKEPKIDFRVYNLLTPDSKVGEVPSVYKRELEEKDGLKEV 780
QY 781 WTSSDDLRNVTWRGADILRGSPTSQTASLGLLTPVSGTQOTKTJRSNPNVEAHLAR 840
Db 781 WTSSDDLRNVTWRGADILRGSPTSQTASLGLLTPVSGTQOTKTJRSNPNVEAHLAR 840
QY 841 PLYGLAVENKGFLOGAPAGEKSGALPOQYPASGENKSKDESQSLRRRGRSGVFCANCL 900
Db 841 PLYGLAVENKGFLOGAPAGEKSGALPOQYPASGENKSKDESQSLRRRGRSGVFCANCL 900
QY 901 TTKTSLWRKNANGGYVCNACGLYQKLHSTPRPLNIIKONNGEQIIRRTKRLNPEALCA 960
Db 901 TTKTSLWRKNANGGYVCNACGLYQKLHSTPRPLNIIKONNGEQIIRRTKRLNPEALCA 960
QY 961 EQLNKOORGSNEBQVNGSPLERRSEDLTESHOREIPLPSLSKYEAQGSILTKSHSAQPV 1020
Db 961 EQLNKOORGSNEBQVNGSPLERRSEDLTESHOREIPLPSLSKYEAQGSILTKSHSAQPV 1020
QY 1021 LVSOTLDIHKWOPLHIQIKSPQESTGDCNSSSVSEGGKSSERGSPIEKYWRPAKHPNY 1080
Db 1021 LVSOTLDIHKWOPLHIQIKSPQESTGDCNSSSVSEGGKSSERGSPIEKYWRPAKHPNY 1080
QY 1081 SPGSPTEIKYQYPLFGLPFWHNDPQSEADWLREWSKYKLSVPCNPHYLHVPLNPPCN 1140
Db 1081 SPGSPTEIKYQYPLFGLPFWHNDPQSEADWLREWSKYKLSVPCNPHYLHVPLNPPCN 1140
QY 1141 YVPYPTFNLPFHSVAVGSONDIPLDIAIKHSRPGPTANGASKEKTAPPNVKNKGPJNV 1200
Db 1141 YVPYPTFNLPFHSVAVGSONDIPLDIAIKHSRPGPTANGASKEKTAPPNVKNKGPJNV 1200
QY 1201 KTEKVDSTRQDELSTKCVHCGIVFLDEVYALHMSCHGSDGPPQCSICQHLCTDKYCF 1260
Db 1201 KTEKVDSTRQDELSTKCVHCGIVFLDEVYALHMSCHGSDGPPQCSICQHLCTDKYCF 1260
QY 1261 HQRGLHRNNAQVEKNGKPKPE 1281
Db 1261 HQRGLHRNNAQVEKNGKPKPE 1281

```

## RESULT 2

```

TRPS_MOUSE
ID TRPS_MOUSE STANDARD; PRT; 1281 AA.
AC Q925H1.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Zinc finger transcription factor Trps1.
GN TRPS1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1].
RZ SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Embryonic gut;
RX MEDLINE=21181829; PubMed=11285235;
RA Malik T.H., Shoichet S.A., Latham P., Kroll T.G., Peters L.L.,
RA Shivdasani R.A.;
RT "Transcriptional repression and developmental functions of the
RT atypical vertebrate GATA protein Trps1.";
RL EMBO J. 20:1715-1725(2001).
CC -!- FUNCTION: transcriptional repressor. May act to restrict
CC expression of GATA-regulated genes at selected sites and stages in
CC vertebrate development.
CC -!- SUBUNIT: Binds specifically to GATA sequences.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: In the embryo, expression is detected in both
CC visceral and skeletal tissues. Found in the maxilla, mandible,
CC snout, prospective phalanges and in the femoral head within the
CC developing hip. Also expressed in the hair follicles.
CC -!- DEVELOPMENTAL STAGE: Detected prior to E7.5, with peak levels at
CC around E11.5. In the developing limbs and face, levels are highest

```

```

CC at E13.5 and decline dramatically thereafter.
CC -!- SIMILARITY: Contains 7 C2H2-type zinc fingers.
CC -!- SIMILARITY: Contains 1 GATA-type zinc finger.
CC
CC This Swiss-Prot entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements
CC or send an e-mail to license@ebi.ac.uk).
CC
CC EMBL: AF346836; AAK39528.1;
CC MGD: MG1:922616; Trps1;
CC InterPro: IPR007087; Znf_C2H2.
CC InterPro: IPR006079; Znf_GATA.
CC Pfam: PF00320; GATA_1;
CC Pfam: PF00396; Zf-C2H2_5;
CC PRINTS: PR00619; GATAZNFINGER.
CC SMART: SM00355; Znf_C2H2_9.
CC SMART: SM00451; Znf_GATA_1;
CC PROSITE: PS00344; GATA_ZN_FINGER_1; 1.
CC PROSITE: PS00114; GATA_ZN_FINGER_2; 1.
CC PROSITE: PS00028; ZINC_FINGER_C2H2_1; 2.
CC PROSITE: PS00157; ZINC_FINGER_C2H2_2; 2.
CC Transcription regulation; Repressor; Zinc-finger; DNA-binding;
CC Nuclear protein; Metal-binding; Repeat.
CC ZN_FING 222 247 C2H2-TYPE 1 (ATYPICAL).
CC FT ZN_FING 333 358 C2H2-TYPE 2 (ATYPICAL).
CC FT ZN_FING 614 637 C2H2-TYPE 3 (ATYPICAL).
CC FT ZN_FING 666 689 C2H2-TYPE 4.
CC FT ZN_FING 692 715 C2H2-TYPE 5.
CC FT ZN_FING 896 920 GATA-TYPE 6.
CC FT ZN_FING 1215 1237 C2H2-TYPE 7.
CC FT ZN_FING 1243 1267 C2H2-TYPE 7.
CC SEQUENCE 1281 AA; 141034 MW; FCA7190C2357898 CRC64;

```

Query Match: 93.8%; Score 6424; DB 1; Length 1281;

Best Local Similarity 93.1%; Pred. No. 0;

Matches 1192; Conservative 39; Mismatches 51; Indels 0; Gaps 0;

```

QY 1 MYRKXNPLRNVAEGEGCILEP-GTSSKVSQKKEFSADOMSENTDQSDAAELNKHKEE 60
DB 1 MYRKXNPLRNVAEGEGCILEP-GTSSKVSQKKEFSADOMSENTDQSDAAELNKHKEE 60
QY 61 SLHVQDPSSSSKKOLKSAVLSEKAGFNYESKGNTPSPFPHDEVTDNMJAFSPAAAG 120
DB 61 SLHVQDPSSSSKKOLKSAVLSEKAGFNYESKGNTPSPFPHDEVTDNMJAFSPAAAG 120
QY 121 VCEPLKSPQRAEADDPQDMACTPSGDSLETKEQKSPKATEETGQAGSGGQANCGLSPV 180
DB 121 VCEPLKSPQRAEADDPQDMACTPSGDSLETKEEFGMGFKATEETGPVQSGGANCGLSPV 180
QY 181 SVASKNPCVPSDGGVRLNKSXTDLLVNDNPPAPLSPPELQDFKNCIGYGYGNDPTDLI 240
DB 181 SVASKNPCVPSDGGVRLNKSXPKGDLVNDNPPAPLSPPELQDFKNCIGYGYGNDPTDLI 240
QY 241 KFKRYHLGLHNRTRQCAELDSKILALHNVYQFSKDFQKYNRSVFSFGLQDINSRPFV 300
DB 241 KFKRYHLGLHNRTRQCAELDSKILALHNVYQFSKDFQKYNRSVFSFGLQDINSRPFV 300
QY 301 LNLGTYDVQVTSGGTF-GIGRKTDCQNTKYFCKFCNFTYMGNSSTELQCHFLQHPN 360
DB 301 LNLGTYDVQVTSGGTF-GIGRKTDCQNTKYFCKFCNFTYMGNSSTELQCHFLQHPN 360
QY 361 KIKASLPSSEVAKPKSEKSNKSIIPALOSSGSDGKWKODKTVKAGDDTPVGYVPIKPL 420
DB 361 KIKASLPSSEVAKPKSEKSNKSIIPALRASGSDGKWKODKTVKAGDDTPVGYVPIKPL 420
QY 421 DSSQMGTEATSYWCKPFCSPSCSSSLKLEHYGKHQFQAVQSGGYNPELNDKLSRGSV 480
DB 421 DSSQMGTEATSYWCKPFCSPSCSSSLKLEHYGKHQFQAVQSGGYNPELNDKLSRGSV 480

```



Db 118 ACDPLKSPIKSEADDTQELASSASVDSLEAKBENDMSPRATDFT--VQCQKVDQSSSPA 175  
Qy 181 SVASKNPQVPSDGGVRLNKSXTDLLVNDNPPAPLSPLELQDFKCNICGYGYGNDPTDCL 240  
Db 176 SVASDNLHVPDSGDIAGLNKSQAVLLVNDNSDAPLSPLELQDFKCNICGYGYGNDPTDCL 235  
Qy 241 KHFRKYHLGLHNRTRQDAELSKLALHNMVQFHSKDFQKNRSVFSGLVDINSRPV 300  
Db 236 KHFRKYHLGLHNRTRQDVELDTKLALHNMVQFHSKDFQKNRSVFSGLVDINSRPV 295  
Qy 301 LLNGYDQVTSGGTFIGIGARTPCQNTKYFRCKFCNFTYMGNSFTELEQHLQTHPN 360  
Db 296 LLNGYDQVTSGGTFIGIGARTPCQNTKYFRCKFCNFTYMGNSFTELEQHLQTHPN 355  
Qy 361 KIKASLPSESAKSEKSNKSIIPALQSSDGLGKWDKITVAKAGDTPVGYGVPIKPL 420  
Db 356 KWKM--SSDSGKPEKSEKSNKSIIPRSCPEGLGKWDKITVAKAGDTPVGYGVPIKPL 412  
Qy 421 DSSRQNGTEATSYWCKFCSCFSCSSSSSLKLEHYGKQHGAVQSGGLNPELNDKLSGSV 480  
Db 413 DSCRQNGTDTNYWCKFCSCFSCSSSSSLKLEHHSKQHGKSGSPNSDUNDEIFRGSV 472  
Qy 481 INQNDLAKSSGEETMTKTDKSSGAKKXDFSSKGAEDNMVSYNCFQCDPRYSKSHQPDV 540  
Db 473 INQNEITSSDEQLPTKDKGL--AKKXDVSSVPTEDI--VTYNCFQCDPRYSKSHQPEV 530  
Qy 541 IVVGPLLRYHQLNHNRKTIKHCPCPRGJCSPEKHGEITYFPACKSKNSCHALLJL 600  
Db 531 ILVGPLLRYHQLNHNRKTIKHCPCPRGJCSPEKHGEITYFPACKSKNSCHALLJL 590  
Qy 601 HSLPCAAGSRVVKHCHOCSETTPDNDVLLPHYESVHESQASDVKQEAHNLQSGDQGSV 660  
Db 591 HSLSGGTSTRVKGHCQDCSCFSSPDVLLHYENAEQAQCEIKQLNHHQHGADGPSI 650  
Qy 661 KESKEHSCTKCDIFIVQEETTSRHYRAHSCVYKRCQCSFATAQTSLEHENTVHCQEQD 720  
Db 651 KEIKEHSCTKCDIFIVQEDLPRIHYRVVNCYKRCQCFNFTAADTOSLHDHNSAHQEF 710  
Qy 721 ITTANGESDGHASTIKKEPKIDPRVYNLLTPDSKMGEPVSSVVKREKLEKQGLKEKV 780  
Db 711 ITTSNGGEH-HGTSISKEPKETDLKVNVLVTPDSKMGAEIIPSTVYKEKLEKDEKLEAK 769  
Qy 781 WTSSSDLLRNVTWKGAILRCSPTQASJGLLTPVSGTQEOYKTLRDSNVAANHLAR 840  
Db 770 WSDGVDLLRGVAMRAPDILRTSPSYQMGGLTIVVNCQKSSRSDSNVAANHLAR 829  
Qy 841 PYGLAVETKGFLOGAPAG--GEKSGALPQOVPASGENKSKDESGLLRRRGGSVFCAN 898  
Db 830 PVYGLSIPKGF-QGVTAGASEKSGQHTPPYPTAGDSKSKDESGLLRRRGGSVFCAN 888  
Qy 899 CLTTKTSLWRKVIANGYVNCAGLYVOKLHSTPRPLNIKQNGEQIIRRTKRLNPEAL 956  
Db 889 CLTTKTSLWRKVIANGYVNCAGLYVOKLHSTPRPLNIKQNGEQIIRRTKRLNPEAL 946  
Qy 959 OAEQLNKOORGNEOVNGSPLERSSEDLHSTESHOREPLSLSKYEAQSGSTKSHSAQC 1018  
Db 949 QPEQUTKQGRASBEQANGSPDITSEDSHSMEGHORENQQLSMKYGSAQSTKSHSAQC 1008  
Qy 1019 PVLVSQTLDIHRMQPLHIQIKSPQESTGDPGNSSVSGKSSRSGSPIEKYWRPAKHP 1078  
Db 1009 TMIVSQTMIDHXMOPHLHIQIKSPQESSGEPGNSSVSGKSSRSGSPIEKYWRPAKHP 1068  
Qy 1079 NYSPPGSPTEKYQYPLFGLPPLVHNFQSEADWLRFWSKYKLSVPGNPHYLSHVPGLNFC 1138  
Db 1069 NYSPPGSPTEKYQYPLFGLPPLVHNFQSEADWLRFWSKYKLSVPGNPHYLSHVPGLNFC 1128  
Qy 1139 QNVYVPTNLPHPHSVAVCSNDIDPLDIAIKHSRGPRTANGASKETKAPVNVKVEGLN 1198  
Db 1129 QNVYVPTNLPHPHSVAVCSNDIDPLDIAIKHSRGPRTANGASKETKAPVNVKVEGLN 1188  
Qy 1199 VVKTEKVDPSQJELSTKCVHCGIVFLDEVMVALHMSCHGDSGPPQCSICQHLCTDKYDF 1258  
Db 1189 VTKIEKSDKSTQJELSTKCVHCGIVFLDEVMVALHMSCHGDSGPPQCSICQHLCTDKYDF 1248

Qy 1259 TTHIQGLHNRNNAQVEKXKPKKE 2881  
Db 1249 TTHIQGLHNRNNAQVEKXKPKKE 1271

RESULT 4  
HUNB\_DRCSE STANDARD: PRT: 757 AA.  
AC C62538;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Hunchback protein.  
GN HB.  
OS Drosophila sechellia (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Lepidoptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
CX NCBI TaxID=7238;  
RN 11;  
RP SEQUENCE FROM N.A.  
RA Tautz D., Nigro L.;  
RT "Microevolutionary divergence pattern of the segmentation gene  
hunchback in Drosophila.";  
RL Mol. Biol. Evol. 15:1403-1411(1998).  
CC -!- FUNCTION: GAP CLASS SEGMENTATION PROTEIN THAT CONTROLS DEVELOPMENT  
OF HEAD STRUCTURES (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE HUNCHBACK FAMILY OF C2H2-TYPE ZINC-  
FINGER PROTEINS.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
DR EMBL; AJ005374; CAA06504.1;  
DR FlyBase; FBcn0024403; Dsec4b.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR Pfam; PF00096; zf-C2H2; 6.  
DR SMART; SM00355; Znf\_C2H2; 6.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 3.  
DR PROSITE; PS01157; ZINC\_FINGER\_C2H2\_2; 2.  
DR Develop. protein; Gap protein; Zinc-finger.  
KW Metal-binding; DNA-binding; Repeat; Nuclear protein.  
FT DOVAIN 61 69 POLY-GLN.  
FT DOVAIN 101 107 POLY-GLN.  
FT DOVAIN 116 121 POLY-HIS.  
FT ZN\_FING 239 261 C2H2-TYPE.  
FT ZN\_FING 269 292 C2H2-TYPE.  
FT ZN\_FING 296 313 C2H2-TYPE.  
FT ZN\_FING 324 348 C2H2-TYPE.  
FT DOVAIN 403 409 POLY-GLN.  
FT DOVAIN 554 562 POLY-GLN.  
FT ZN\_FING 704 726 C2H2-TYPE.  
FT ZN\_FING 732 756 C2H2-TYPE.  
SQ SEQUENCE 757 AA; 31846458029550E CRC64;

Query Match 4.0%; Score 273.5; DB 1; Length 757;  
Best Local Similarity 19.3%; Pred. No. 3.9e-08;  
Matches 173; Conservative 113; Mismatches 346; Indels 263; Gaps 32;

Qy 472 NCKLSGSGVINGKCLAKSSGEGMTWTKDSSCAKKKDFSSKGAEDNMV7-----SYNQC 526  
Db 26 NIKQEGCHLDGNSVASSPRQSPFSTNHLQELKQCCQHQCCQPMDTJCACTPSPQNDQ 85  
Qy 527 FCFRYSKSGSPGVIVVGLLHHV-----QQLNKHKCTIKHCFPCPRG;CSPEKH----- 577  
Db 86 NSLCHYDNLQQLLQCCQYQHQHFAAQCCQHHHMLMGCFNPLTPPGLFNPQHFGYGN 145



```

QY 578 LGBITPFAKCKNCSCALLLLHLSGAAGSRVKGHCQCFTTPDVLPHFYESVH 637
Db 146 LRPSPOTPTASTIAPVAV-----ATGSS-----EKLQALTPMDVT--PKPSA 189
QY 638 ESQASDVQKQANHLQSGDGOOSVK---ESKEH----- 666
Db 190 KSSQSNIEPEKHEQDMSNSEDKMYAESDEDDTNRIMPYNSEGRKMKYKCTCGVVA 249
QY 667 -----SCTKCDFTQVBEISRHYR--RAHSCVKKQCQCSFTAAATQ 705
Db 250 TKVDFWAHTRMKPKDILQCPKCPFFTEPKHLEHYIRKHQKQFPQCDKCSYTCVNS 329
QY 706 SLLEHFN-----VHCQSQDIT-----ANGEEDGHAISTIKER-----PKIFRVY 747
Db 310 MLKSHRASHSVYQYRCACDQYATKYCHSPKFLHRKYGRKPGVMVDEDTGTFNLSVIDV 369
QY 748 NLLT-POSKMGPEVSESVVREKLEEKDGLKERVWTESSDCLNVTWRCADILAGSPSY 826
Db 370 GTRGPKSKNGGFIASG----- 866
QY 807 TOASLGLLTVPVSGTQETKLRDPSNPVYEAHLARPIYGLAVETKGFQAGAPGEXSGAL 866
Db 392 SRKS-----NVAAVAPQOQSQPAQPVATSHLSAALQGFPL-----VQSNAPPAASPVJ 441
QY 867 PQQYPASGENK--SKDESQSLRRRRSGGVFCANCLTTSKLRKXNANGVYCNACGLYQ 924
Db 442 P--LPASPAKSVASVEQTPSLP-----ANLLPPLASLLQQRNMAFF-----PYW 486
QY 925 KLHSTPRPLNIQNGNEQIIRTRKRLNPEALQAEQLNKQOQSGNEEQVNGSP:ERRS 984
Db 487 NLN-----LQMLAQQAQAVL-----ACLSPRVREQLQQNQOQSDNEEDQDDEVERKS 536
QY 985 EDHLTESHQREIPLPSLSKYAEAGSLTKSHSAQPVLVLSOTLDHKMQLPH:QIKSPQE 1044
Db 537 VDSAMD-----SQGTPVKEDBQQQ-----QQQFLAXNLKVEEB 571
QY 1045 STGDPGNSVSEBQSGSERGSPTEKYMRAKFNHNSPPGSPTEKYQYPLFGLPFVHND 1104
Db 572 AT-PLMSSNASRRKG---RVKLDTLLQ-----L 597
QY 1105 QSEADMLRFKSKYKLSVPGNPHYLHVPLNPNQNVYPTNLPHFSAVGSNDNDPL 1164
Db 598 RSEA-----MTSPBOLKVPSTP-----MPTASSPIAGKPM-----EBHCSGTSADE--- 64:
QY 1165 DLAIKH-----SRPGPTANGASKETKAPPNVKNKGPVNVVYKTEKYDRSTQDE 1212
Db 642 SMERPHVRQANTSASTASSGNSNASSNSNGSSSSNGTTSAAVAPPSGTAAAG 701
QY 1213 LSTKCVHGVFLDEVYALHMSCHGSGPQCSICQHLCTDKYDFTTHIORGLH 1267
Db 702 AYECKYCDIFFKDAVLYTIHMGYHSCDDVFKCMCGEKCDGPGVGLFVHMARNAH 756

RESULT 5
HUNB_DROME STANDARD; PRT; 758 AA.
AC P05084; Q24018;
DT 01-APR-1988 (Rel. 07, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hunchback protein.
GN HB CR CG9786.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R; TISSUE=Embryo;
RA Tautz D., Lehmann R., Schuerch H., Schuh R., Seifert E.,
RA Kienlin A., Jones K., Jaekle H.;

```

```

RT "Finger protein of novel structure encoded by hunchback, a second
RL member of the gap class of Drosophila segmentation genes.";
RN Nature 327:383-389(1987);
RP REVISION TC 525.
RA Tautz D.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX MEDLINE=94259232; PubMed=8203478;
RA Margolis J.S., Borowsky V., Shim C.W., Posakony J.K.;
RT "A small region surrounding the distal promoter of the hunchback gene
RL directs maternal expression.";
RN Dev. Biol. 163:381-388(1994);
RP SEQUENCE FROM K.A.
RC STRAIN=Berkeley;
RX MEDLINE=20396506; PubMed=10731132;
RA Adams M.C., Geiniker S.B., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaral-Andes P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vardell M.D., Zhang C., Chen L.X.,
RA Brandon S.C., Rogers Y.-H.C., Blazer V., Chao M., Pfeiffer B.D.,
RA Abril J.F., Adaya A., An H.-J., Andrews-Flannock C., Baldwin D.,
RA Ballaw R.M., Basu A., Baxendale C., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bermet B.P., Bhattacharya D., Bolshakov S.,
RA Brkova D., Botchan N.R., Bouck J., Brockstein P., Brettler P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry C.M., Cawley S., Dahlke C., Davenport J.B., Davies P.,
RA De Fabrics B., DeCher A., Deng Z., Mays A.D., Dew I., Dietz S.X.,
RA Dedson K.J., Doolittle L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Petraz C., Ferreira S., Fleischmann W.,
RA Finkler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Goddek A., Gong P., Gorrell J.H., Gu Z., Guo P., Harris M.,
RA Harris N.L., Harvey D., Heiman T., Hernandez J.S., Houck J.,
RA Hoskins R., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaalali M., Kalish F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.E., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mares B., McIntosh T.C., McLeod X.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris C., Moshrefi A.,
RA Mount S.G., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.B.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazzone M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheer F., Shen H.,
RA Shue B.C., Sider-Kimons I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sur E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Kassam A.D., Weissstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye C., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RT "The genome sequence of Drosophila melanogaster.";
RN Science 287:2185-2195(2003);
RP SEQUENCE OF 281-349 FROM N.A.
RX MEDLINE=93066327; PubMed=1438276;
RA Sommer R.J., Retzlaff M., Goerlich K., Sander K., Tautz D.;
RT "Evolutionary conservation pattern of zinc-finger domains of
RL Drosophila segmentation genes.";
RN Proc. Natl. Acad. Sci. U.S.A. 89:10782-10786(1992);
RP POLYMERIZATION.
RA Tautz D., Nigro J.;
RT "Microevolutionary divergence pattern of the segmentation gene
RL hunchback in Drosophila.";
RN Mol. Biol. Evol. 15:1403-1411(1998);
CC -!- FUNCTION: GAP CLASS SEGMENTATION PROTEIN THAT CONTROLS DEVELOPMENT
CC OF HEAD STRUCTURES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: IN EMBRYO. EXPRESSION OF MATERNAL TRANSCRIPT

```



RX MEDLINE=2238257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Stachenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein X.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Lequellain N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.K.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettenan M., Madan A.C., Rodrigues S., Sanchez A.,  
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Brakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield V.S.N., Krywinski M., Skalska U., Smalus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RA "Generation and initial analysis of more than 15,000 full-length  
 R human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- FUNCTION: BINDS AND ACTIVATES THE ENHANCER (DELTA-A ELEMENT) OF  
 CC THE CD3-DELTA GENE. FUNCTIONS IN THE SPECIFICATION AND THE  
 CC MATURATION OF THE T LYMPHOCYTE. ALSO INTERACTS WITH A CRITICAL  
 CC CONTROL ELEMENT IN THE TBT (TERMINAL DEOXYNUCLEOTIDYLTRANSFERASE;  
 CC PROMOTOR AS WELL AS WITH THE PROMOTERS FOR OTHER GENES EXPRESSED  
 CC DURING EARLY STAGES OF B AND T CELL DEVELOPMENT.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=7;  
 CC Name=Ik1;  
 CC IsoId=Q13422-1; Sequence=Displayed;  
 CC Name=Ik2;  
 CC IsoId=Q13422-2; Sequence=VSP\_006848;  
 CC Name=Ik3;  
 CC IsoId=Q13422-3; Sequence=VSP\_006850;  
 CC Name=Ik4;  
 CC IsoId=Q13422-4; Sequence=VSP\_006847, VSP\_006850;  
 CC Name=Ik5;  
 CC IsoId=Q13422-5; Sequence=VSP\_006852;  
 CC Name=Ik6;  
 CC IsoId=Q13422-6; Sequence=VSP\_006849;  
 CC Name=Ik7;  
 CC IsoId=Q13422-7; Sequence=VSP\_006851;  
 CC -!- TISSUE SPECIFICITY: ABUNDANTLY EXPRESSED IN THYMUS, SPLEEN AND  
 CC PERIPHERAL BLOOD LEUKOCYTES AND LYMPH NODES. LOWER EXPRESSION IN  
 CC BONE MARROW AND SMALL INTESTINE.  
 CC -!- SIMILARITY: BELONGS TO THE IKAROS FAMILY OF ZINC-FINGER PROTEINS.  
 CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;  
 CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/Ikaros1258.html".  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC  
 CC EMBL: U40462; AAC50459.1; -;  
 CC DR EMBL: S80876; AAB50693.1; -;  
 CC DR EMBL: BC018349; AAB50693.1; -;  
 CC DR HSP: F15822; IBB0.  
 CC TRANSFAC: T02702; -;  
 CC Genew: HGNC:13176; ZNF1A1.  
 CC MIM: 603023; -;  
 CC GO: GO:0003677; F:DNA binding activity; TAS.  
 CC GO: GO:0007498; P:mesoderm development; TAS.  
 CC InterPro: IPR007087; Znf C2H2.  
 CC Pfam: PF00096; zf-C2H2; 5.  
 CC ProDom: PD000003; Znf C2H2; 2.  
 CC SMART: SM00355; Znf\_C2H2; 6.

DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 5.  
 DR PROSITE; PS0057; ZINC\_FINGER\_C2H2\_2; 4.  
 KW Transcription regulation; Activator; Zinc-finger; Metal-binding;  
 KW DNA-binding; Nuclear protein; Repeat; Alternative splicing.  
 FT ZN\_FING 117 139 C2H2-TYPE 1.  
 FT ZN\_FING 145 167 C2H2-TYPE 2.  
 FT ZN\_FING 173 195 C2H2-TYPE 3.  
 FT ZN\_FING 201 224 C2H2-TYPE 4.  
 FT ZN\_FING 462 484 C2H2-TYPE 5.  
 FT ZN\_FING 492 514 C2H2-TYPE 6.  
 FT VARSP-IC 10 53 Missing (in isoform Ik4).  
 FT VARSP-IC 54 140 Missing (in isoform Ik2).  
 FT VARSP-IC 54 283 Missing (in isoform Ik6).  
 FT VARSP-IC 197 283 Missing (in isoform Ik3 and isoform Ik4).  
 FT VARSP-IC 197 238 Missing (in isoform Ik7).  
 FT VARSP-IC 141 283 Missing (in isoform Ik5).  
 FT CONFLICT 11 12 QV -> PS (IN REF. 2).  
 FT CONFLICT 214 214 S -> T (IN REF. 2).  
 FT CONFLICT 245 245 N -> X (IN REF. 2).  
 FT CONFLICT 296 296 MISSING (IN REF. 2).  
 FT CONFLICT 298 298 S -> T (IN REF. 2).  
 FT CONFLICT 352 355 KPLA -> RRS (IN REF. 2).  
 FT CONFLICT 372 372 N -> Y (IN REF. 2).  
 FT CONFLICT 420 426 PHARNGCL -> RRAQRY (IN REF. 2).  
 FT SEQUENCE 519 AA; 57528 MW; 7B0129C4E3F41A8 CRC64;  
 Query Match 3.9%; Score 269; DB 1; Length 519;  
 Best Local Similarity 25.9%; Pred. No. 4; e-08;  
 Matches 139; Conservative 67; Mismatches 192; Indels 138; Gaps 29;  
 QY 614 LTPVSGTQQTITLRESNPVAAHARPIYGSVAVTKFQLQAPA---GGEKSGALPQY 87C  
 DB 37 LSTTSGQGSSK---SDRVASN-----VKVETQDEENGACRACNGBEACEDRLX 95  
 QY 671 PASGENKS---KDESCLLRRRGSSV-----FCA-NCLTTTSLWKNANG 913  
 DB 96 DASGEKNGSHRQGGSSAL-----SGVGIR-PNGLKCDICGICIGPNVVMVHRSH 140  
 QY 914 G---YVCNACGL-----YQKLHSTPRPLNIKQNGEQIIPRRTRKLNPEALQA 960  
 DB 141 GERPFQCNCCGASFTCKGNLLRHILHSGEKP---FKCHLCNYACRRR-----DALTG 190  
 QY 961 EQLNKQGRSGNEEQVNGSPLEERS--EDHLTESHQ--REIPLPS-----LSKYEAQGS---1009  
 DB 192 ELRTHSVGKPHKCGYGRSYKQRSLEBHKRCHRYLESMLGPTLYPVYKKEETHSEMA 250  
 QY 1010 --LTKSHSACQPVLSQTLDIHKRMQPLHIQKSPQESTGDPGNSSSVSEKSGSSERGSP 1067  
 DB 251 ELCLCKTGSERSLVLORLASNVAKRYSK-----PQKELGKGLSDTPYDSSASEKENE 304  
 QY 1068 IEK-----YVRPAKHPNVSPG-----SPIEKYQYPLF-GLPFPV 1100  
 DB 305 MKKSHVMDCAINNAIYLGAEISLRP-VQ---TPPGGSEVVPVISDMYQLHKPLAECTFRS 361  
 QY 1101 HKDFQSEA-DWLRFWSKYYKLSVFGNPHVLSHVFGPL-PQCQNYVPYPTFNLPPHPSAVGS 1158  
 DB 362 NHAQDSAVENLLJLSKAKL-VF-----SEREASPSNSCQDSTTESNNEEQRSGLIYL 414  
 QY 1159 DNDIPLDLAIKHSRPGPTANGAS-XEKTKA-----NGLSLKEERAYDLLRAAENSQDARVVST-----SGSQM 459  
 DB 415 TNNI-----APRAR-----NGLSLKEERAYDLLRAAENSQDARVVST-----SGSQM 459  
 QY 1213 LSTKQVHGIVFJDEVYVALHMSCHGDSPPFCCSICQHLCTDKYDFTTHIQRLHR 1268  
 DB 460 KYVKECHRVLF-LDHVYV-TIEMGCHGFRPPFCNCGYHSDRYEFSHH-IRGEHR 515

```

RESULT: 7
HUNB_DROOR          STANDARD;          PRT; 767 AA.
ID   HUNB_DROOR
AC   O62537;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Hunchback protein.
GN   HB.
OS   Drosophila orena (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7233;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Tautz D., Nigro L.;
RT   "Microevolutionary divergence pattern of the segmentation gene
RT   hunchback in drosophila.";
RL   Mol. Biol. Evol. 15:1403-1411(1998).
CC   -!- FUNCTION: GAP CLASS SEGMENTATION PROTEIN THAT CONTROLS DEVELOPMENT
CC   OF HEAD STRUCTURES (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE HUNCHBACK FAMILY OF C2H2-TYPE ZINC-
CC   FINGER PROTEINS.
CC
CC   This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC
CC   EMBL; AJ005375; CAAG6505.1; --
CC   FlyBase; FBgn0024421; Dore\bb.
CC   InterPro; IPR007087; Znf.C2H2.
CC   Pfam; PF00096; zf-C2H2; 6.
CC   SMART; SM00355; Znf.C2H2; 6.
CC   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
CC   PROSITE; PS0157; ZINC_FINGER_C2H2_2; 2.
CC   Developmental protein; Gap protein; Zinc-finger;
KW   Metal-binding; DNA-binding; Repeat; Nuclear protein.
FT   DOMAIN          61..71
FT   DOMAIN          104..110
FT   DOMAIN          119..124
FT   ZN_FING         242..264
FT   ZN_FING         271..293
FT   ZN_FING         299..321
FT   ZN_FING         327..351
FT   ZN_FING         410..415
FT   DOMAIN          530..533
FT   DOMAIN          686..693
FT   DOMAIN          698..703
FT   ZN_FING         714..736
FT   ZN_FING         742..766
SQ   SEQUENCE 767 AA; 84207 MW; 9C7710B7AFED78914 CRC64;

Query Match          3.9%; Score 266.5; DB 1; Length 767;
Best Local Similarity 19.0%; Pred. No. 9.7e-08;
Matches 170; Conservative 116; Mismatches 354; Indels 257; Gaps 30;

QY 472 NKLRSGVINQNDLAKSSEGETMTYTKSSGAKKDFSSKGAEDNMVT-----SY 523
D 26 NIKOEPGHLGDNVASSPROSPIPSTNHLEQFLKQQOQHQQOQPMDTLCAMTPSPFSQ 85
QY 524 NCOFCDFRYSKSGHPDIVVUGPLLRHY---QQLNIHKCTTGKCPFCPRGLCSPEKH-- 577
D 86 NDQNSLQHYDANLQQGLLQQOQQOQQHQAQQOHHHHHLMGFGNPLTPPG:PNPMQHFY 145
QY 578 ---LGEITVPFACRKNCSHCALLLLHLSPGAGSGSRVKEQCHQCSFTTPDQVLLFHYE 634
D 146 GGNVLRSPQPTPISASTVASVAV-----ATGSS-----EKLOALTTPPKCVT--PPK 189

```

```

QY 635 SVHESQASDVYKQFAHRLQSGDGGQSVK---ESKEH----- 666
D 190 SPAKSSQSNIEPEKEHEDQMSNSEDKMYMVESEDDDTNIRMP-YNSHGKMKVYKCKTCGV 249
QY 667 -----SCTKCDPITQVEEISRHYR--RAHSCYKQCQCSTAA 702
D 250 VAITKYDFMAHTRHMKDPDKILOCPKCPFVTEFKHLEHYH-RKHKKQKPFQCDKCYTCV 309
QY 703 DTCSLLHFHNT-----VHCQEDDIT-----ANGEEDGHAISTIKER-----PKIDF 744
D 310 NKSMKLSHRKSHSVYQYRCADYATKYCHSKLHRLKRYCHKPGMWLDEGTPNSLVI 369
QY 745 RYVNLCT-PCKYGEVPVSQVYKLEEKGLKEXVWTESSDQJRNVTWRGADILRGS 833
D 370 DVTGTRRGPKSKAGGPIASG-----GS-----GS 393
QY 804 PSYTOASGLLTQVSGTQEQTKT.RDSPNVAAHARPIYGLAVEYKGF--LOGAPAGGE 861
D 394 GSGSGSRKSNVAAPVQCCQCTQPTOP-----P*SLCAAJGGFPLVOSNAPPA 442
QY 862 KSGALPCQYPASGENSKDESQSLRRRRRGSGVFCANCLTTKTSJMRKNANGGVVNCACG 921
D 443 ASPLLPLPVSFPAKSVASVEQTPSLRSP-----ANKLPLLASLLQONHMAFF----- 489
QY 922 LYOKLHSTPRLNIIKONNGEQI:RRTRKRLNPEALQAEQANKQORGSEEQVNGSPLE 981
D 490 FYMNLK-----LQMLAAQQQAVAL-----AQLSPRYRECLQQNQCCSDNEEEQDEYE 539
QY 982 RRSDDHUTESHQREIP:PSLSKYAEAGSLTKSHSAQPVLSVOTLCHKE--MQPLHIQI 1039
D 540 RKSVDSSAMCL-----SQGTPVKEEQO-----LHCQOQOQLVMNL 576
QY 1040 KSPQESTGDFGNSSSVSEKSGSSRSP:EKYMRPAKHPKPNYSPGSP:EKYQYFLEGLPF 1099
D 577 KVEEAT-PLVSSSNARRKG---RVLKJLDTLLQLRSGVWTSPE----- 616
QY 1100 VHNDFOSEAEWLRFSWKYKLSVFGNPHYLSHVPLGNPFCNQYVYPTFNLPFHFSAVGSD 1159
D 617 -----QLKVFSTP-----MPTASSPIAGRKPM-----EDHCSGTSSA 649
QY 1160 NDIPDLAIKH-----SRPGPTANGASKETKAPPNVKNEGPLNVYKTKVDASTQ 1210
D 650 DESNETAHRVCANTSSASTASSGNSNASSNGSSSSSSSSSSSSSSSSSSSSSSSSSSSS 1267
QY 1211 DELSTKCHVCGIVFJBEWYALHMSCHGDSGPFQCSICQHCCTKCYDFTTHIQRLH 1267
D 710 AGA:VECKYCTIFFKDAVLVTHMGVHSCDVFKNCMGCKGCDGPGVGLFVHVARNAI 766

RESULT 8
Y296_HUMAN          STANDARD;          PRT; 1829 AA.
ID   Y296_HUMAN
AC   O15015;
DT   16-OCT-2001 (Rel. 42, Created)
DT   16-OCT-2001 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Hypothetical zinc finger protein KIAA0296.
GN   KIAA0296.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Nagase T., Ishikawa K.-I., Nakajima D., Chira M., Seki N.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. VII.
RT   The complete sequences of 133 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
DNA Res. 4:141-150(1997).

```

CC -!- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -!- SIMILARITY: BELONGS TO THE KRUPEL FAMILY OF C2H2-TYPE ZINC-  
CC FINGER PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AB002294; BAA20756.1; -;  
DR InterPro; IPR007087; Znf C2H2.  
DR Pfam; PF00096; Zf-C2H2; 28.  
DR SMART; SMO0355; Znf C2H2; 31.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 30.  
DR PROSITE; PS0057; ZINC\_FINGER\_C2H2\_2; 29.  
KW Hypothetical protein; Transcription regulation; DNA-binding;  
KW Zinc-finger; Metal-binding; Nuclear protein; Repeat.  
FT ZN FING 8 31  
FT C2H2-TYPE.  
FT ZN FING 48 70  
FT C2H2-TYPE.  
FT ZN FING 75 97  
FT C2H2-TYPE.  
FT ZN FING 239 261  
FT C2H2-TYPE.  
FT ZN FING 266 288  
FT C2H2-TYPE.  
FT ZN FING 294 316  
FT C2H2-TYPE.  
FT ZN FING 374 396  
FT C2H2-TYPE.  
FT ZN FING 401 424  
FT C2H2-TYPE.  
FT ZN FING 465 487  
FT C2H2-TYPE.  
FT ZN FING 492 514  
FT C2H2-TYPE.  
FT ZN FING 575 597  
FT C2H2-TYPE.  
FT ZN FING 617 639  
FT C2H2-TYPE.  
FT ZN FING 644 666  
FT C2H2-TYPE.  
FT ZN FING 821 843  
FT C2H2-TYPE.  
FT ZN FING 848 870  
FT C2H2-TYPE.  
FT ZN FING 881 904  
FT C2H2-TYPE.  
FT ZN FING 958 980  
FT C2H2-TYPE.  
FT ZN FING 1052 1074  
FT C2H2-TYPE.  
FT ZN FING 1079 1101  
FT C2H2-TYPE.  
FT ZN FING 1203 1225  
FT C2H2-TYPE.  
FT ZN FING 1230 1252  
FT C2H2-TYPE.  
FT ZN FING 1258 1280  
FT C2H2-TYPE.  
FT ZN FING 1299 1321  
FT C2H2-TYPE.  
FT ZN FING 1326 1348  
FT C2H2-TYPE.  
FT ZN FING 1364 1386  
FT C2H2-TYPE.  
FT ZN FING 1557 1579  
FT C2H2-TYPE.  
FT ZN FING 1585 1607  
FT C2H2-TYPE.  
FT ZN FING 1677 1699  
FT C2H2-TYPE.  
FT ZN FING 1704 1726  
FT C2H2-TYPE.  
FT ZN FING 1732 1754  
FT C2H2-TYPE.  
FT ZN FING 1761 1783  
FT C2H2-TYPE.  
SQ SEQUENCE 1829 AA; 200760 MW; B20C482B38684895 CRC64;

Query Match 3.7%; Score 251.5; DB 1; Length 1829;  
Best Local Similarity 18.4%; Pred. No. 2.3e-06;  
Matches 279; Conservative 159; Mismatches 485; Indels 593; Gaps 77;

QY 56 HKEHSLHVQDPSSSSKKD-LKSAVLSEKAGFNYESPSKGGNFFSPHDEVTDRNMLA-- 112  
DB 26 HRE--LLHPSRQDSEADSIIPRYCQCQCGYRHPGSLN-----HRRTHETGLFPCT 78  
QY 113 -----FSPPAAGVCVEPLKSPORAEADD-----PQDMACTPS-----GSGSJE 149  
DB 79 TCGKDFSNPMA-----LKSHMRTHAPEGRRRHRPRPKEATPHLQGETVSTQGWQRLG 132  
QY 150 TKEDQKMSKATEETCQAQS-----GCAN-----CQGLSPVSV 182  
DB 133 SSEGMENQTKTEETPDCESPDPRAAGTWMEDLPTROREGSLASHPGPEDGADGMPSTN 192  
QY 183 ASKNPQVPSDGGVRLN-----KSKTLLVNDNPDPAFLSPQLQDFKNCICG--YGVYG 233  
DB 193 SARAPLPPIPASSLLSNLEQYLAEVSNVFTGQGEPTQSPFAEERRRYKSCGCKTYRHAG 252

QY 234 NPDPTOLIKFRKYHLGLHNRTQDAELDSKILALHNMYQFSHSDKDFOKVNRSVFSGVLQD 293  
DB 253 S-----LTHNRQSHTLGIYPCACFKPEF-SNLMALKN-----HSR----- 286  
QY 294 INSSRPVLLNGTYQVQVSSGTFIGIRKTPCQCQNTKYFRCKFCNFTYMGNSSTELQCH 353  
DB 287 -----LHAGY-----RPHCPHCPRVF--RLPELLEH 312  
QY 354 FLOTH-----ENKIKASLPSEVAKPKSEKSNKSI-PALQS-----SDSGELGKWQCK 400  
DB 313 -CQSHEGEBQERWEKMPPTIN--GHIDEQQDLPSAQLMNGSAELSTGEL----- 363  
QY 401 ITVKAGDQTPVGVSVPIKPLDSSR-----ONGTEATSYMYCKFCFSCESSSXX 450  
DB 364 -----EDSGLEBYRPFRCGDCGRTYRHAGSLNHRKSHQGVYPCSLCSKGLFNAALK 417  
QY 451 LLEHYGKQHGAVCGSLNPELNCKLSRGSVINQNDLAKSSEGETNKTCKSSSGAKKQDF 510  
DB 419 --NVRARHRRPQGVGENGQ--PSYPPAPLLLAETTHKEEDPTTLCRR----- 463  
QY 511 SSKGAEDNMVSYNQCFDFRYSKSGSDVIVUGPLLRHYQCLNHRKCTIKKCFPCERG 570  
DB 464 -----PYKSECGRAYRHR-----GSLVNH-----RSHRTGEVCCSCJCR- 499  
QY 571 LSSPEKHLGEITYP--FACRKNCSNCHC-----ALLLHLSPGAAGSSRYK 613  
DB 500 -----KYPNLXALRNHVYHCKAARRSACI-GAEGAFSHLKVLEFPDPVEAERP 549  
QY 614 H--QCHQCSFTTPOVTVLLPHYESVHESQASVYKAEHQLQSSDQOQSVKESKEHSCYK 671  
DB 549 HTDQCHVCK-----HEEATDITPAAD-----KTAAHICISIC 580  
QY 672 DFTQVEEESIRH-----YRAHS-----CYKROCSTAAATQSLLEFNTVH 715  
DB 581 GULLFEDASSLERHGLTHGAGEKENSRTEITMPPAPACRDCGKSYRHSGSLNHRCT-H 539  
QY 716 QCEQCITTANGEEGHAISTIKKEPKIDFRVYNLITPQSKMGEVSVSVKREKJEEKDG 775  
DB 640 -QTGDFSCGACAKPHMTMAAK-----NHLRRHSR-----RRSRHRKRA 678  
QY 776 -----LKERVWTSSSDDLRNTWRGADILRGSPTSQASLGLTTPVSGTQPCPT 824  
DB 679 GGASGGREAKLLAESWTRELD-----NEGLES----- 708  
QY 825 KTLRDSPNVEAAFLARPIYGLAVETKGLQAGAGKSGALPQQYPASGENKSKDESOS 884  
DB 709 ----QDPSSGESPHGAG-----NLESDGDLQAESEGCCKGLERDETHFGDKESGTCBG 760  
QY 885 LLRR-----RRSGV-FC-----ANCLTTTSLWR 908  
DB 761 LERKDAJLDNLDIPGEEGGTHFCDSLTGVDGQKPKATQPNSSSSHSANAVTG-----MQ 816  
QY 909 KWANGOVVNCAGLYOKLHSTPRP-NIL-----KQNGEQIIRRRTRK 951  
DB 817 --AGAAHTGSDCG-----HSFPHTAGLLSHRCHPPPGIYCCSLCPKFPDLSALASHFON 869  
QY 952 RLNPALQAE-----QLNKQ-----RGSNEE-QWNG----- 977  
DB 870 HRPGEATSACPFCCCLCGMIFPGRAGYLRHRQAHSSSGTMTEGSEEGEEGVAAEAPAR 929  
QY 978 SPLRRSEDLTHSHOREI-PLPSL-----SKYEAQSLTKSHSAQCPVLVSQTL 1026  
DB 930 SPFLQJSEAEJLNQLOREVEA-LSAGYGHICCCGCTYDDLGLSLRRHHQSQ-----SSGT 984  
QY 1027 J-HKRNQPLHI-----CIKSPCESTGDP-GNSSSVSESGK 1060  
DB 985 TACKAPSLGVAGAGANEXYVDSVLEDI-VNSVSEGGDAKS-QEGAGTPUGDSJC-CQGE- 1042  
QY 1061 SBERGSFIKKYVRPAK-----HFNYSPPGSPIEKYCPFLFGLPFVHNCFOSEACWL----- 1111  
DB 1043 ----SLLEAQAPRPFRCNCGKTYRHGSLN-----HRKIHQTCFLCPVCS 1085

```

QY 1112 RFWXKSLVPCNPHYLHVPLGPNPCQVYPTENLPPHFSAVGSDNPLDJAIAKHS 1171
DQ 1086 RCY-----PFLAAYRNHLRHPR-CRGSF-----QVCPIDEAAGSS-----ELCV--- 1125
QY 1172 RPTGANGASKETKAPPNVKNKGP-----LAV-----VET 1202
DQ 1126 ---GPIPEGSNK-----PQMAEEGQGAQVEKQLQELKVEPLEVARVKEEWETTAKG 1179
QY 1203 EKV-----DRSTODELSTK-----CVHCGIVFLDEVMYALHMSCHGSGPFGQCSICQHL 1251
DQ 1180 EIEIPRLETAEGCQTEASSRPFSCVCGRSYKHAAGSLINRQSH-CTGHFGCCACSKG 1236
QY 1252 CTDKYDFTHIQRGHLH 1267
DQ 1239 FSNLMLXNH---RRTH 1252

RESULT 9
HUNB_DROYA
ID HUNB_DROYA STANDARD; PRT: 759 AA.
AC O62541;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update);
DE Hunchback protein.
GN HB.
OS Drosophila yakuba (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7245;
RN [1]
SEQUENCE FROM N.A.
RA Tautz D., Nigro L.;
RT "Microevolutionary divergence pattern of the segmentation gene
hunchback in Drosophila.";
RL Mol. Biol. Evol. 15:1403-1411(1998).
CC -!- FUNCTION: GAP CLASS SEGMENTATION PROTEIN THAT CONTROLS DEVELOPMENT
OF HEAD STRUCTURES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: BELONGS TO THE HUNCHBACK FAMILY OF C2H2-TYPE ZINC-
FINGER PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AJ005376; CA06506.1; -
DR FlyBase; FBgn0022824; Dyak\hb.
DR InterPro; IPR037087; Znf C2H2.
DR Pfam; PF00096; zf-C2H2; 6.
DR SMART; SM00355; ZNF_C2H2; 6.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 2.
KW Developmental protein; Gap protein; Zinc-finger;
KW Metal-binding; DNA-binding; Repeat; Nuclear protein.
FT DOMAIN 61 70 POLY-GLN.
FT DOMAIN 103 109 POLY-GLN.
FT DOMAIN 118 123 POLY-HIS.
FT ZN_FING 241 263 C2H2-TYPE.
FT ZN_FING 270 292 C2H2-TYPE.
FT ZN_FING 298 320 C2H2-TYPE.
FT ZN_FING 326 350 C2H2-TYPE.
FT DOMAIN 405 410 POLY-GLN.
FT DOMAIN 525 528 POLY-GLU.
FT DOMAIN 555 564 POLY-GLN.
FT DOMAIN 690 695 POLY-ALA.
FT ZN_FING 706 728 C2H2-TYPE.
FT ZN_FING 734 758 C2H2-TYPE.

```

```

SQ SEQUENCE 759 AA; 93287 MW; 67D37A252978D365 CRC64;
Query Match
Best Local Similarity 19.3%; Score 248.5; DB 1; Length 759;
Matches 174; Conservative 11; Mismatches 344; Indels 271; Gaps 36;

QY 472 NDKLSRGSVINQNDLAKSSEGETYNTDKSSGAKKCFSSKGAEDNMVT-----SY 523
DQ 26 NIKCEPGHEDGNSVASSERQSPISPTNHLCEFLKQQQ-QQQHQQQPMDTLCATVSPSQ 84
QY 524 NQCFCDFRYSKSGPDIIVVGPILLHY-----CQJHNIHKCTIKHCPCPCRGCLSPK 577
DQ 55 NQONSICHYDASIQQLQQQLQQQYQCHQFAAQQQHHHHHLMGCFNPLTPGLPNMWHFY 144
QY 578 ---LGEITYFACRKGNCSHCALLLHLSPAGASGRVZKHQCQCFSTPDVDVLLFHYE 634
DQ 145 GGNLRPSQPTPTASTVAPVAV-----ATGSS-----EKLCAITFPMDVT---PPK 188
QY 635 SVHESQASDVQKQANHLQSGDQGSVK-----ESKEH----- 666
DQ 189 SPAKSSQSN-EPEKEHCHQMSNSEDMMKYMAESEDDTNIRYP-VNSHGKVKYKCKTCGV 248
QY 667 -----SCCKCDPITOVEEELSIRHYR-RAHSCVKCEQCSPTAA 752
DQ 249 VAKYDFWATHTHMKPKILQCPKCPVTEFKHLHSHIRKHKQKPKFCDCRCSFTCV 328
QY 703 DTOSLLEHFN-----VHCQEQDIT-----ANGEEDGHAISTIKKE-----BKIDF 744
DQ 309 NKSVLNSHRKSHSVYCYRCACQDQATKYCHSFKLHLRYGKPGMVLEDTGTPNPSLVI 368
QY 745 RVYNLLT-PDSXGEPVSESVVREKLEEDGKLEKWTESSDDLRNVTWRGADILRGS 803
DQ 369 DVYGTTRGPKSKNGGPIASG-----GSGSGS-----RKNVAAVA 423
QY 804 PSYTOASLGLLTPVS-GTQCTKTLRDSNVE---AAHLARPIYGLAVETKGLQAGAPAG 859
DQ 404 PQCCQSQ-----PAQPATLSLSALQGPFLVQNSAPPAASPVLP-----PAS 447
QY 860 GEKSGALPQOVPASGKNKSKDESQSLRRRGRSGVFCANCLITKTSLWRKNANGGVYCN 919
DQ 448 PAKSVASVEQTTPSLPSP-----ANLLPPLASLQQRNRYAFF---484
QY 920 CGLYQKLHSTFRPNJ::KQNGEQIIRRTKRLNFEALQAEOLNKCQSGSNEEQVNGSP 979
DQ 485 ---PYWNLN-----QMLLAQQAQAAVJ-----AQLSPRMESQLCCQCHOSDNEEBEQDDE 532
QY 980 LERPSEDLHTESHQREIPLPSLSKYCAQGLSKSHSAQOPVLVSQTL-D-HKRMCPHLHIQI 1039
DQ 533 YERKSVDSAVQLSQ-GTPVKKEDDQHQQQ-----QQP-----CQPLAKNL 570
QY 1040 KSPQESTGDFGNSVSVSEKSGSSERSPIEKYVRPAKHNPSPGSPIEKYQYPLGLPF 1099
DQ 571 KVEEAT-PLMSSSNASRRKG---RVLKLTDTLQ----- 600
QY 1100 VHNDFQSEADWLRFASKYKLSVPGNPHYLHVPLGPNPCQVYPTENLPPHFSAVGSD 1159
DQ 601 ---LXSEA-----MTSPCEKLKVPSTP-----MPTASSPIAGRKPMP-----EDHCSQTSSA 643
QY 1160 ND-PLDJAIAKH-----SRPGPTANGASKETKAPPNVKNKGPVNVKTEKVD 1207
DQ 644 DE---SMETHAVACANT-SASSTASSSGNSGNASSNGNSS---NSSNGTSSAAAAPASGT 699
QY 1208 STODELSTKCVHCGIVFLDEVMYALHMSCHGSGPFGQCSICQHLCTDKYDFTHIQRGHL 1267
DQ 699 PAAAGAIYECKYCDIIFKQAVLYT-HWGVHSCDEVPKCYCGEKCDGPGVGLFVHARNAH 758

RESULT 10
TCF8-CHICK
ID TCF8-CHICK STANDARD; PRT: 1114 AA.
AC P36137; C42408;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)

```

DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Transcription factor 8 (Delta-crystallin enhancer binding factor)  
DE (Delta EFL).  
OS Gallus gallus (Chicken).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN (1)\_TaxID=9031;  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lens;  
RX MEDLINE=9411644; PubMed=7904558;  
RA Fuchasini J.-I., Sekido R., Murai K., Kamachi Y., Kondoh H.;  
RT "Delta-crystallin enhancer binding protein delta 3FL is a zinc  
RT finger-homodomain protein implicated in postgastrulation  
RT embryogenesis.";  
RL Development 119:433-446(1993).  
RN (2)  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo;  
RX MEDLINE=97082972; PubMed=8964504;  
RA Sekido R., Takagi T., Okanami M., Moribe H., Yamamura M., Higashi Y.,  
RA Kondoh H.;  
RT "Organization of the gene encoding transcriptional repressor delta3FL  
RT and cross-species conservation of its domains.";  
RL Gene 173:227-232(1996).  
CC -!- FUNCTION: BINDS TO DELTA 1-CRYSTALLIN ENHANCER CORE AND REPRESSSES  
CC LENS-SPECIFIC TRANSCRIPTION. IT BINDS AS WELL MANY OTHER NON-LENS  
CC SPECIFIC DNA SEQUENCES.  
CC -!- SUBUNIT: MAY INTERACT WITH CTBP1 (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -!- TISSUE SPECIFICITY: EXPRESSION IS DEVELOPMENTALLY REGULATED WITH  
CC HIGH EXPRESSION IN MESODERM, NERVOUS SYSTEM AND LENS.  
CC -!- DEVELOPMENTAL STAGE: EXPRESSION STARTS AFTER GASTRULATION, WHEN  
CC ORGANOGENESIS HAS JUST BEGUN.  
CC -!- SIMILARITY: BELONGS TO DELTA-5FL/2FH-1 FAMILY OF TWO-HANDED ZINC  
CC FINGER/HOMEDOMAIN PROTEINS.  
CC  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: D14313; BAA03259.1; -  
CC EMBL: D76434; BAA1178.1; -  
CC EMBL: D76433; BAA1178.1; JOINED.  
CC FIR: I50222; I50222.  
CC TRANSFAC: T01467; -  
CC InterPro: IPR001356; Homeobox.  
CC InterPro: IPR007087; Znf\_C2H2.  
CC Pfam: PF000096; zf\_C2H2; 7.  
CC ProDom: PD000010; Homeobox; 1.  
CC ProDom: PD000003; Znf\_C2H2; 2.  
CC SMART: SM00389; HOX; 1.  
CC SMART: SM00355; Znf\_C2H2; 7.  
CC PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 5.  
CC PROSITE: PS50157; ZINC\_FINGER\_C2H2\_2; 7.  
CC Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;  
CC Homeobox; Metal-binding; Repeat.  
CC ZN\_FING 170 193 C2H2-TYPE.  
CC ZN\_FING 200 222 C2H2-TYPE.  
CC ZN\_FING 240 262 C2H2-TYPE.  
CC ZN\_FING 268 292 C2H2-TYPE (ATYPICAL).  
CC ZN\_BIND 581 640 HOMEBOX-LIKE.  
CC SITE 767 771 CTBP-BINDING MOTIF (BY SIMILARITY).  
CC ZN\_FING 904 926 C2H2-TYPE.  
CC ZN\_FING 932 954 C2H2-TYPE.  
CC ZN\_FING 960 981 C2H2-TYPE (ATYPICAL).  
CC DOMAIN 989 1081 GLU-RICH (ACIDIC).  
CC CONFLICT 299 299 K -> N (IN REF. 2).

SQ SEQUENCE 11-4 AA; 123:50 MW; D914284143E7D273 CRC64;  
Query Match: 3.6%; Score 248.5; DB 1; Length 11-4;  
Sest Local Similarity 18.8%; Pred. No. 1.7e-06;  
Matches 210; Conservative 144; Mismatches 377; Indels 385; Gaps 52;  
QY 378 NENKSTIPALQSSDSGLGKW---CKKITVAG-----DTPVGVSVPIKPLDSRQN 426  
DB 22 NYNNVISA--NSDSDSDKHLHVEESITDAACDASVPEDLPTCHTV--LPENSEEG 77  
QY 427 GTEATSYWCKFCFSCESSSLKLEHYGKHGAVQSGGLNPELNTKLSGVSINQDL 486  
DB 78 STN-----SC-----WEDEGKETKEL---LQPE----- 97  
QY 487 AKSSEGETWTKCKSSSAKKKDPSSKGAELNMVTSYNCQCFDPRYSKSHGPDVWVGP 546  
DB 98 ASDDEVGCTVKEDECDCAENSHQNPVEFL-----QOEDTAVIYPE 141  
QY 547 LRYVQQLNHNHKT-----IKHCPFCPRG---LCSPEKHLGEITYEFACRKS 590  
DB 142 APEEDQRCQTPASQDENGTEPAFSQLITCPYCDRGYKFTSLKER---KYRHEKNEJ 198  
QY 591 NNSHCALLLHLSPGAUSSRYKHQCHOCSTTPVQVLLFHYESVHESQASDVK--QEA 646  
DB 199 NFS-----CSLCSYT-----FAYRTQLDRHMTSHKSGRQJ 228  
QY 649 NHLQSGDGGQSYKESKESCTKCFITQVEEEISRYHRAHS---CYKCROCSFTAACTP 705  
DB 229 RAVTQSGNRKFK-----CTCGKAFYKXHLKELHJ--RHSGEXPYECPCCKRFSHSG 281  
QY 706 SLLEHNTVHCEQDITTANGEEGHAISTIKKEPKIDFRVYNLTTP--DSKMGEPVSES 763  
DB 282 SYSSHSSKKC-----GLMPVKRAESGLKTSQCSPPSLSPSPSPARPQ 327  
QY 764 VVKREKLEKCKLKE---KVWTE-----SSDDLNVNVRGADILGSS 803  
DB 328 I--RQXIKNKP--LCEQGFVNQIKTEPVDFYEFKPIVVASGINCSTPLQGVFGSGPLQAT 384  
QY 804 PS---YTQA---SLGLLTPVS-----GTQEQTKT 826  
DB 385 SFGQGVQAVVFTVGLVSPISNLSDQNVLVKAVDGVNIPQVLENNHNLASKEQ-ET 443  
QY 827 LRDSPNVNAAH-----LAPPIVGLAVETKGFQGPAPAGEKSGALPQQYF 871  
DB 444 SNAS--QAGHSLISALSPLVDQDGTTHIINYSLEQPSQLQVVPQNLKKEHSV?--- 499  
QY 872 ASGENKSKDESQSLRRRRSGVFCANCLTTKTSLWRXQANGG-----YVCNACGLYQK 925  
DB 500 ---TNSCKNEK-----LPEDLTVSEK--DRNFEGETNDSTCLLCOOC----- 537  
QY 926 LHSTPRPLNITKONGEQIIRRTTRKRLNPEALQALNKOORGSGNEEQVNSPLRSE 985  
DB 538 ---PGDLNALQE-----LKHVEK--NPPOL-----POSSSGTEAEKPPSPAPSETGE 579  
QY 986 DHLTESHQREIPSP--LSKYEAQGSILTKSHSQQPLVLSQT-----LDIHK-----RMCPH 1036  
DB 580 NNLSPG---QPPLKLLSLKAYALNAQPSAELSILADSINLPLDVVVKWFWFKMAQAG 636  
QY 1037 IQKKS-----PQESTGDCNCSNVSEK-----GSSER 1064  
DB 637 ISVCSSGSPSPQVKISSPTDNDQDAA--TNESEPQNSNTNSCNPA--TSSQSSGGSTQN 696  
QY 1065 GSPIEKYMRPAKHP-----NVSPPGSPTEKIQYPL--FGLPFVHNDFOSEAOWL 1111  
DB 697 GS---RSSTPSPSPNLSSRSNSQGYTYAEGVQBPQWEPDLQSLPKQHGSEJLERSTIT 753  
QY 1112 PFWSKYKLSVFGNPHYL-----SHVFG--PNPCQNYVYPTENLPE 1151  
DB 754 SVYQNSVYVQVEEFLNLTCAKKEPKQKNSITDSDPIVNVIPPSANPINAIPTVTAQSLP 813  
QY 1152 HPSAVGSDNDIP-----LDLAIKHSREGPT--ANGASKK-- 1184  
DB 814 -IVAIAGQNSVPCLRALANKQTLIPQVAVTYTSTVSPAVQETPPKQTCANSGSCERQD 872

```

QY 1185 --TKAPPNVKNEPLNVVYKTEKVDSTODELSTKCVHCGIVFELDVVYALHMSCHGDSGP 1242
DB 873 TSSEGVSNVEDONDSDSTPPKKKRRKTENGMYA-CDLCKIFQKSSLLRHKKVEHTKRP 93:
QY 1243 FQCSICQHLCTDKYDFTTHIORGLHRNNAQVEKNGK 1278
DB 932 HECGICKAFKHKHKLHIEHMLHSEKPYQCDKCGK 967

RESULT 11
ABFL_HUMAN
ID ABFL_HUMAN STANDARD; PRT; 3703 AA.
AC Q15911; O15101; Q13719;
DT 16-OCT-2001; (Rel. 40, Last sequence update)
DT 15-SEP-2003; (Rel. 42, Last annotation update)
DE Alpha-fetoprotein enhancer binding protein (AT motif-binding factor)
DE (AT-binding transcription factor 1).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A. (ISOFORM A).
RC TISSUE=Lung;
RX MEDLINE=96070776; PubMed=7592926;
RA Miura Y., Tam T., Ido A., Morinaga T., Miki T., Hashimoto T.,
RA Tamaoki T.;
RT "Cloning and characterization of an ABFL1 isoform that expresses in a
RT neuronal differentiation-dependent manner.";
RL J. Biol. Chem. 270:26840-26848(1995).
RN [2]
SEQUENCE FROM N.A. (ISOFORM B).
RC TISSUE=Hepatoma;
RX MEDLINE=92049333; PubMed=1719379;
RA Morinaga T., Yasuda H., Higashio K., Tamaoki T.;
RT "A human alpha-fetoprotein enhancer-binding protein, ABFL1, contains
RT four homeodomains and seventeen zinc fingers.";
RL Mol. Cell. Biol. 11:6041-6049(1991).
RN [3]
SEQUENCE OF 1-1190 FROM N.A. (ISOFORM A).
RX MEDLINE=99425270; PubMed=10493829;
RA Loftus B.J., Kim U.-J., Shedden V.P., Kalush F., Brandon R.,
RA Fuhrmann J., Mason T., Crosby M.L., Barnstead M., Cronin S.,
RA Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,
RA Eichler E.E., Harris P.C., Venter J.C., Adams M.D.;
RT "Genome duplications and other features in 12 Mb of DNA sequence from
RT human chromosome 16p and 16q.";
RL Genomics 60:295-308(1999).
RN [4]
SEQUENCE OF 1151-3703 FROM N.A.
RA Kozlowski A., McQuerry Y., Hotic M.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Transcriptional activator that binds to the AT-rich core
CC sequence of the enhancer element of the AFP gene.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=A;
CC Name=B;
CC IsoId=Q15911-1; Sequence=Displayed;
CC IsoId=Q15911-2; Sequence=VSP_006825;
CC -!- SIMILARITY: Contains 4 homeobox domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL: L32832; AAC14462.1; -
CC EMBL: D10250; BAAC1095.1; -
CC EMBL: AC002044; AAC31674.1; -
CC EMBL: AC004943; AAC79153.1; -
CC HSP: P20263; 10CP.
CC TRANSFAC: T00049; -
CC TRANSFAC: T01665; -
CC Genes: HGNC:777; ATBF1.
CC MIM: 104155; -
CC GO: GO:0005634; C:nucleus; TAS.
CC GO: GO:003705; F:RNA polymerase II transcription factor acti...; TAS.
CC GO: GO:006355; P:regulation of transcription, DNA-dependent; TAS.
CC InterPro: IPR001356; Homeobox.
CC InterPro: IPR007087; Znf-C2H2.
CC Pfam: PFC00046; homeobox_4.
CC Pfam: PF00096; zf-C2H2_18.
CC PROSITE: PSC0010; Homeobox_4.
CC PROSITE: PSC0027; Homeobox_1; 2.
CC PROSITE: PSC0071; Homeobox_2; 4.
CC PROSITE: PSC0028; ZINC_FINGER_C2H2_1; 14.
CC PROSITE: PS50157; ZINC_FINGER_C2H2_2; 9.
CC Transcription regulation: Activator; Zinc-finger; Metal-binding;
CC DNA-binding; Homeobox; Nuclear protein; Repeat; Polymorphism;
CC Alternative splicing.
CC ZN_FING 282 305 C2H2-TYPE.
CC ZN_FING 640 663 C2H2-TYPE.
CC ZN_FING 671 694 C2H2-TYPE.
CC ZN_FING 726 750 C2H2-TYPE.
CC ZN_FING 804 828 C2H2-TYPE (ATYPICAL).
CC ZN_FING 945 968 C2H2-TYPE (DEGENERATE).
CC ZN_FING 984 1009 C2H2-TYPE (ATYPICAL).
CC ZN_FING 1040 1064 C2H2-TYPE (ATYPICAL).
CC ZN_FING 1088 1112 C2H2-TYPE (ATYPICAL).
CC ZN_FING 1223 1246 C2H2-TYPE (ATYPICAL).
CC ZN_FING 1252 1275 C2H2-TYPE.
CC ZN_FING 1360 1385 C2H2-TYPE.
CC ZN_FING 1401 1423 C2H2-TYPE.
CC ZN_FING 1429 1452 C2H2-TYPE.
CC ZN_FING 1545 1569 C2H2-TYPE.
CC ZN_FING 1596 1620 C2H2-TYPE.
CC ZN_FING 1993 2006 C2H2-TYPE.
CC ZN_BIND 2145 2204 HOMEBOX_1.
CC ZN_BIND 2428 2501 HOMEBOX_2.
CC ZN_FING 2328 2552 C2H2-TYPE (ATYPICAL).
CC ZN_FING 2530 2552 C2H2-TYPE.
CC ZN_FING 2641 2700 HOMEBOX_3.
CC ZN_FING 2711 2734 C2H2-TYPE.
CC ZN_BIND 2944 3003 HOMEBOX_4.
CC ZN_FING 3024 3048 C2H2-TYPE.
CC ZN_FING 3529 3553 C2H2-TYPE.
CC DOMAIN 104 107 POLY-PRO.
CC DOMAIN 460 489 POLY-GLU.
CC DOMAIN 770 784 POLY-ALA.
CC DOMAIN 1723 1743 POLY-GLN.
CC DOMAIN 1789 1794 POLY-GLN.
CC DOMAIN 1852 1857 POLY-GLN.
CC DOMAIN 2337 2352 POLY-PRO.
CC DOMAIN 3197 3209 POLY-GLN.
CC DOMAIN 3210 3214 POLY-PRO.
CC DOMAIN 3231 3231 POLY-GLN.
CC DOMAIN 3376 3389 POLY-GLN.
CC DOMAIN 3392 3395 POLY-GLN.
CC DOMAIN 3507 3527 POLY-GLY.
CC DOMAIN 3597 3600 POLY-PRO.
CC DOMAIN 3636 3639 POLY-SER.
CC VARSPL-C 1 914 Missing (in isoform B).
CC /FTId=VSP_006825.
CC A -> V
CC /FTId=VAR_011694.
CC MISSING.
CC /FTId=VAR_011695.
CC G -> GGG.

```



```

FT CONFLICT 72 72 /FTID=VAR_011696
FT CONFLICT 422 422 A -> P (IN REF. 3).
FT CONFLICT 579 579 A -> S (IN REF. 3).
FT CONFLICT 767 767 I -> S (IN REF. 3).
FT CONFLICT 777 777 A -> V (IN REF. 3).
FT CONFLICT 846 849 HRRV -> RHLG (IN REF. 3).
FT CONFLICT 997 997 A -> S (IN REF. 3).
FT CONFLICT 1150 1190 EEAIEVEGSETPADPERLAKDQEGGASSOAEKELTSP
FT -> GEWSHRGRPRGLGVHLLSTSRGLDFEGVYCPAGPH
FT VPY (IN REF. 3).
SQ SEQUENCE 3703 AA; 404468 MW; 0562AF37D4DCF856 CRC64;

Query Match 3.6%; Score 245; DB 1; Length 3703;
Best Local Similarity 17.8%; Pred. No. 1.4e-05;
Matches 292; Conservative 214; Mismatches 515; Indels 620; Gaps 80;

QY 22 EPIGTESKVGKNEFSADQMSNTQSDAAELNHKEE-----HSLHV----- 64
DB 575 EGVRTNVAEGRRLLDF-ADP-SANKONATAPERNESTEGDGGFVPHHGHAGSLCELGVG 632
QY 65 QDPSSSKDK-K-SAVISEKAGNYESPCKGNP-----SFPHDEV 105
DB 633 ECPSSGVECPKCDTVLG-----SSRLGGHMTVYHSRNSCKTLKCPKCNHXYKQOT 685
QY 106 TDRNMLAFSPAAAGVCEPLKS-----PQRAEAD-----DPQDMACTPSCD-SLE 145
DB 686 LEAHMKE-KHPEPGGCVYCKSQPHPR-LARGSYTCGVKPKFRCEVCNVSTTKGNLSIH 744
QY 150 TKEDQMSKATETQOAGSQANCOG--LSPVSVASKNPQVPDSGVGLNKSCTELLVN 237
DB 745 MQSDKLNHNQNLQNGGEGVFHTAGAAAAAANAANSSSCGA----- 792
QY 208 DNEPAPLSPLO-DPKNICIGVYNGNDPDLKHFR-----KVHLG----- 249
DB 793 ----PSTKTKPTWRCEVDY-----ETVARNLRIMHSEKHMNMMLQGNMTQI 842
QY 250 LHNRT-----QDAELDSKILA-----LHNMVQFHSKDPQVNRSVFSGVLCD----- 293
DB 843 QNHHRVLGSLPAPABAELOYVLAQNMNLPNLMKDSAAADQ-----FMWSGFQDPA 898
QY 294 INSSRPVLLNGTYDVQVTSG-----TFIGIGR-----KTPDCQGNTKYFRCPFN-FT--- 341
DB 899 MAATPALVGEIPLDMRLGGGLVSEELMNLGSEFIQND--PSLXFFQCAVCKRFTD 956
QY 342 ----YNGNS-----STELQHFLOTHPNKIKASLPSEV 371
DB 957 NLDMLGLHMNVERSLSEDEKAVMGDSYQCKLCRYNTQLKANP-QLHCK----- 1004
QY 372 AKPSEKSNKSI-PALOSSDSGLKQWDKITVAGDDTFVGYSVPIK-----PLDS 422
DB 1005 ---TDRHVQYQLVAHTEGGEKANERLKCVA-----IGNPVHLKCNACDYNTSLSEK 1054
QY 423 SR-----QNGTEATS-YVWCKFCGSCSSSSSLKLEHY-GKH 459
DB 1055 LR.LHTVNSRHEASLKLYHLQCHESGVEGSCYVHCVLCNYS---TKAKLNL-CHVRSMKH 1112
QY 460 GAVQS-----GGLNPELNDKLSRGSVINQ-----NDLAK 488
DB 1113 QRSESLRKLORLOKGL-PEDEDLQGTFTIRRCPTSTOPBEA-EDVEGPSETAADPEELAK 1171
QY 489 SSEG---ETWTKTDKSSSGAKKDFSSKGAEDNMVTS-----YKCOFCQFR 531
DB 1172 DQEGGASSQAELTDSPTSRIKSPGSSSEPLSKSPKPAEE-KPEQMYQCYPC--K 1229
QY 532 YKSHGPDVIVGWPLLRHYQLQNLNHHKCTIKHGPFPCPRGLCSPKXHLGEITYPFACRKN 591
DB 1230 YSNAD-----VNLRLVHANTQHSVQ---PMLRCP-LCCDM-----NW 1263
QY 592 CSHCALLLHLSPGAGSSRVKHQCHQ---CSTTDPDVV--L.LPHVESVHESQASDVQK 646
DB 1264 KHLQLHQLTHL-----HSVAPDCVEKLIMTVTPEMVPSMFLPAAVPPDRGKSNLE 1316

```

```

QY 647 EA-----NH.LQSDGQO-----SVKSEKHSCTX--CDFITQVEEE 680
DB 1317 EACKQSETEDLGKN:LPSASTEQSGOLKPSPADPGSVREDSGFIWKRKGNQVPTSA 1376
QY 681 ISRHVRAHS-----CYKRCQCSPT-----AAAT----- 704
DB 1377 LQTHFNEVAKRPQLEFVSRHVKVYACNCCSLAFKTIKQLHSHQYHVRANTMCLCQR 1436
QY 705 -----GSLLEHNTVHCQ---EODI-----TTANBEEQHA:STIKSEPKIDFRVYNLLT 751
DB 1437 SFTTFQALKKHLETSHLE:SEADIQOGLLANGD-----LLA 1475
QY 752 PDSKGEPV---SESVVXREKLEEKDGKLEK-----VWTESSSDDLNENVWRCAG 798
DB 1476 -----MGDFLAEDCHIIIEEDKEESDLEDKSPGSSSGSVQEDSGSEPRKALPFR--- 1529
QY 799 LDRG-----SPSYTQASLGLL---TPVSGTQFOITKLRDSENVFAA 836
DB 1529 --KGNPTMEKP-LPSEPVKCTVCKESFTQKILLVHYNSVSHJHLKRALQESATGOPE 1586
QY 837 HLARP-----YQ-----LAVETKGLQAPAGGEXSGALPQOYPASGENKSKD 880
DB 1587 PTDSPONKPFKNTONVAYSQSSTLEIHXPVSLHQTAKAAKLEA-----ASGSSNGTS 1640
QY 861 ESQSL-LRRRRSGGVPCANCLTKTSLMRKNANGGVVNCAGLYQKLLHS-----TP 930
DB 1641 KSSSISLSSTSTSPVSTSGSNTFTTS---NPSSAGIAPESNLLSOVPTEVGMPLGNP 1696
QY 931 KPLNLT-KQNGNFCITRRTRTKRLNFEA:CAEOJNKQCGSNEECVNGSPLERRS---EDHL 988
DB 1697 IGAKEASPSEPREANRKKLADVIASRQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 1756
QY 989 TESHORELP-----PSLSKYEAQGSLLTKSHSACQPVLV-----SOTLDIHKR--- 1031
DB 1757 QSLQOQCALIQSOLENPTLLPHFPXT:ETLLOQOQOHLFPFVIPAHEFCNLNEVSLP 1816
QY 1032 -----YQPH:QIKSPQES-----TQPCNSSSVSEKSGSSRSGSPLEKY 1071
DB 1817 VTSGALT:GTGPGLEDCUKAQVQVPOQSHQO:LPOQOQNCLSLAQSHSA-----L 1867
QY 1072 MRAPKPNVSPGSPLEKYQYPLGLPVHDFQSEADLRFWSKYKLSVPCNPVHLSHV 1131
DB 1868 LQPSQHPKKNKLVIKK-----EKESQERD----- 1894
QY 1132 PGLNPNQNVVPYPTFNJPHPSAVGSDNDI-----PLDLAIKHSPGPTANGASKERTKA 1187
DB 1895 -----SAGEGEGNTGPKETLPDALKAKEKKE:APGGSGSPSVL 1932
QY 1188 PPNVKNEGP:NVVKT-----EKVDRSTQDELSTK--CVHCGVVF 1224
DB 1933 PPIASDARGNATKALLENFGFELVIQVNNENKQVCKXNGKTQDGENLEKUECDSCGKLF 1992
QY 1225 LDEWYVALHYS-CHGSDSGPFO 1244
DB 1993 SNLLILKSHQEHVHQVYPPFK 2013

RESULT 12
GAT2 XENLA
ID -GAT2 XENLA STANDARD; PRT: 452 AA.
AC P23770;
DT 01-NOV-1991 (Ref. 20, Created);
DT 01-NOV-1991 (Ref. 20, Last sequence update);
DT 16-OCT-2001 (Ref. 40, Last annotation update);
DE Transcription factor xGATA-2 (GATA binding factor-2);
GN GATA-2;
OS Xenopus laevis (African clawed frog);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xeropodidae; Xenopus.
CX NCBI TaxID:5355;
RN 1;
RP SEQUENCE FROM N.A.

```

Xenopodinae; Xenopus.  
 KCB1\_TaxID=8355;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX XEDLINE=88052679; PubMed=2826129;  
 RA Raiz A, Altshul A., Perry O, Keefe H., Melton D.A.;  
 RT "Xfin: an embryonic gene encoding a multifingered protein in  
 RT Xenopus.";  
 RL EMBL J. 6:3365-3370(1987).  
 RN [2]  
 RN CHARACTERIZATION.  
 SP MEDLINE=94321366; PubMed=7492399;  
 RX Andreazoli X., de Lucchini S., Costa M., Barsacchi G.;  
 RA "RNA binding properties and evolutionary conservation of the Xenopus  
 RT multifinger protein Xfin.";  
 RL Nucleic Acids Res. 21:4218-4225(1993).  
 RN [3]  
 RN STRUCTURE BY NMR OF FINGER 11.  
 RX XEDLINE=89346749; PubMed=2503871;  
 RA Lee M.S., Gippert G.P., Scher A.V., Case D.A., Wright P.E.;  
 RT "Three-dimensional solution structure of a single zinc finger DNA-  
 RT binding domain.";  
 RL Science 245:635-637(1989).  
 RN [4]  
 RN STRUCTURE BY NMR OF A FINGER.  
 RX MEDLINE=89178224; PubMed=2506074;  
 RA Lee M.S., Cavanagh J., Wright P.E.;  
 RT "Complete assignment of the 1H NMR spectrum of a synthetic zinc  
 RT finger from Xfin. Sequential resonance assignments and secondary  
 RT structure.";  
 RL FEBS Lett. 254:159-164(1989).  
 CC [1]- FUNCTION: BINDS RNA. COULD FUNCTION IN POST-TRANSLATIONAL  
 CC REGULATION PROCESSES.  
 CC [1]- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC [1]- TISSUE SPECIFICITY: COCYTES AND IN SPECIALIZED CELL TYPES SUCH  
 CC AS NEURAL RETINA CONES.  
 CC [1]- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT OOGENESIS AND  
 CC EMBRYOGENESIS.  
 CC [1]- DOMAIN: CONTAINS 37 FINGER MOTIFS IN 6 DOMAINS.  
 CC [1]- PTX: PHOSPHORYLATED.  
 CC [1]- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
 CC FINGER PROTEINS.  
 CC [1]- SIMILARITY: Contains 1 KRAB domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sb.ch](mailto:license@sib-sb.ch)).  
 CC -----  
 CC EYBL: X06321; CAA29425.1;  
 CC RDB: ZNF; 15-OCT-91;  
 CC InterPro: IPR01905; KRAB.  
 CC InterPro: IPR07087; Znf\_C2H2.  
 CC Pfam: PF03352; KRAB; 1.  
 CC Pfam: PF00096; zf-C2H2; 36.  
 CC ProDom: PD000003; Znf\_C2H2; 20.  
 CC SMART: SM00349; KRAB\_1  
 CC SMART: SM03355; ZNF\_C2H2; 35.  
 CC PROSITE: PS00605; KRAB; 1.  
 CC PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 35.  
 CC PROSITE: PS00157; ZINC\_FINGER\_C2H2\_2; 37.  
 CC Zinc-finger; Metal-binding; RNA-binding; Repeat; 3D-structure;  
 CC Phosphorylation.  
 CC FT LCMAIN . 56 KRAB.  
 FT ZN\_FING 108 130 C2H2-TYPE.  
 FT ZN\_FING 136 159 C2H2-TYPE.  
 FT ZN\_FING 164 186 C2H2-TYPE.  
 FT ZN\_FING 192 214 C2H2-TYPE.  
 FT ZN\_FING 233 242 C2H2-TYPE.  
 FT ZN\_FING 256 265 C2H2-TYPE.

```

FT ZN_FING 248 270 C2H2-TYPE.
FT ZN_FING 276 298 C2H2-TYPE.
FT ZN_FING 326 348 C2H2-TYPE.
FT ZN_FING 354 376 C2H2-TYPE.
FT ZN_FING 382 404 C2H2-TYPE.
FT ZN_FING 410 432 C2H2-TYPE.
FT ZN_FING 438 460 C2H2-TYPE.
FT ZN_FING 466 488 C2H2-TYPE.
FT ZN_FING 503 525 C2H2-TYPE.
FT ZN_FING 531 553 C2H2-TYPE.
FT ZN_FING 559 581 C2H2-TYPE.
FT ZN_FING 587 609 C2H2-TYPE.
FT ZN_FING 615 637 C2H2-TYPE.
FT ZN_FING 643 665 C2H2-TYPE.
FT ZN_FING 671 693 C2H2-TYPE.
FT ZN_FING 699 721 C2H2-TYPE.
FT ZN_FING 750 772 C2H2-TYPE.
FT ZN_FING 778 800 C2H2-TYPE.
FT ZN_FING 806 828 C2H2-TYPE.
FT ZN_FING 834 856 C2H2-TYPE.
FT ZN_FING 862 884 C2H2-TYPE.
FT ZN_FING 890 912 C2H2-TYPE.
FT ZN_FING 918 940 C2H2-TYPE.
FT ZN_FING 988 1010 C2H2-TYPE.
FT ZN_FING 1016 1038 C2H2-TYPE.
FT ZN_FING 1044 1066 C2H2-TYPE.
FT ZN_FING 1136 1158 C2H2-TYPE.
FT ZN_FING 1164 1186 C2H2-TYPE.
FT ZN_FING 1192 1214 C2H2-TYPE.
FT ZN_FING 1220 1242 C2H2-TYPE.
FT ZN_FING 1248 1270 C2H2-TYPE.
FT ZN_FING 1276 1298 C2H2-TYPE.
FT STRAND 1045 1045
FT STRAND 1052 1052
FT HELIX 1056 1062
FT HELIX 1063 1065
FT HELIX 1066 1066
SQ SEQUENCE 1350 AA; 27F10AB0851E0AD8 CRC64;

Query Match
Best Local Similarity 19.2%; Score 238; DB 1; Length 1350;
Matches 239; Conservative 137; Mismatches 390; Indels 476; Gaps 70;

QY 222 PKNICGGYGNDDTLIKHFRVYHGLNHRTRQDAELDSKILALHNMVQF-----273
DB 354 YLCSHCNKGFIQ--SDLVKHFT-HTG--ERPQCAECHKGFIQKSDLVKHLATH-GEK 408
QY 274 ---SHSKDFOKVNRVSFGLVDINSSRPVLLNCTYDVQVTSQGTFGIGRKTPTDC---326
DB 409 PFKSHC-DKXFTERSALAKHQRHTGTGKPY-----KSCDCKE 446
QY 327 -----QNTKYFRCKFCNFTYMGNSSTLEQHFLOTHPNKIKASLPSSVAK 373
DB 447 FTQRNLILHQRHTGERPYKTLCDRTFIQNS--DLVKH-----QKVHANPLSD---495
QY 374 PSEKNS-----NKSIPALOS-----SDSGDLGKWDKITV 403
DB 496 PHTANSPHKSKCOLTFGHWSTFMKSHLHSGEKKFQCAECKGFTQKSDLVK---HIRV 552
QY 404 KAGDDTPGVSVPIKPL---DSSRNGTQEARSYW-----CKFCSPSCSSSLK-LJ 452
DB 553 HTGEK-----PFKLLCKKSFSQNSD--LHKHWRHTGKPPCYPCYCDKSFERSALI 603
QY 453 EHYGQHGAVOSGGLNPELNDKLSRGSVINQNDLAKSSEGTMTKTKDSSGSAKKKDFSS 512
DB 604 KHRHTHTG-----ERPHKCSVCQKGFQKSA---LKHSEHTHTGEX-----642
QY 513 XGAEDNMVTSYNCQCDPRYSKSGHPDVIVGGLRHVQQLHNTHK---CT-----560
DB 642 -----PYPTQC-----GKSPIQNSDLVKH-QR-HTGKPPVHCTECKKRTGEG 683
QY 561 ---LKH-----CPFCRGLCSPE---KH-----GE-----ITYPFAKRN--591

```

```

DB 684 SSLVXKRRTHSGEKPYRCPCQCEKTFIQSSDLVXHLVWANGENPPAATAFHEILIRENLT 743
QY 592 -----CSHCA-----LALLLSPGAGSSRVKQCHCCS---FTTQDVLLEHY 633
DB 744 RSEPDPYFCTEGCVPHCARPALLKHL---RTHKTEKYPCHNECDKSPQTQSD---LVKH 796
QY 634 ESVHESQASDVKQEAN--HLQSGSQQSVKESKEH-----SCTKCD--FITQVEEELSR 683
DB 792 LRTH-GERPYHCPENCKGFIQNSD---LVKHQPTHGERPYTCSCODKGFQIR--SALT 851
QY 684 HYR--RANSHCYKRCQCSFTADTQSLLEHPTVHQCQODITTANGEEDGHAISTIKERPK 742
DB 852 HNRCTHTGKPYKCECCQKCFQNSDLVKH-QR-H-----TGEKPYHC-----892
QY 742 IDPRVYNLTTPSKMGEPPVSESVKREKLEK-----DGLKEKVTWESSDDLNRVWTKG 796
DB 893 -----PCKDKRFTGESSLKHQRIHSRIKPYPCGVCKGKSFQSS-----931
QY 797 ADILRGSPSYTOASLGLTTPVSGTQE-----QTKTRDSNPVAAHLARPIYGLAVETKG 851
DB 932 -NLKELKXCHSEQN-----PPVALSSELGFAETQTHDPDVJ---H1---VVG---DTAS 976
QY 852 FLOGAPAGGKSKSALPOQYPASGENKSKDECSLLRRRAGSGVFCANCLTTKSLWRKNA 911
DB 977 YI-SPEAAGERS-----LYOKLHSTPPPLNIKKONNGEOTIRRTKRKLNPEALCAE 961
QY 988 ---FKNDCGCKCFARHSV---KXVR-HTGERPYK-----SOCTRSFIQXS 1029
QY 962 QLNKQORNSQNSQNGSELEERS--EDHLTJESHQREIPLPSLSKYEAGSUTKSHSAQOP 1019
DB 1030 DLVKHYR-HTGERPYKGLCERSFVEKSGALRRHQ-----VHKNESE 1071
QY 1020 VLVSQTLDIHKEMQPLHLCIKSPQESTGDPGNSSSVSECKSSBERGSP--LEKYNRPKHX 1077
DB 1072 VL-----NSAME--QCCVTYWGSKDDP--NSLVQPLHVIKEESPHIVNAYSPLSIL 1120
QY 1078 PNYSP-----GSPLEKYYPLFLPFVHNDQSEADMLRFWSKYKLSVPGN-----1124
DB 1121 QSYFPILPEPKGP--RYSCEGCKCFHRSV-----FLKXHWHTGTGOPYTCKEGKSGF 1173
QY 1125 -----PHYLSHVGLPNPCQNYVPYPTFNLPHFSAVGSNDCLPLDLA-KHSPGPTA 1177
DB 1174 SSSALVKVRIHTGKPYPC-----STGKSFIOKSDLA-KHQR-----1213
QY 1178 NGASKEKTKAPENVKNEGLNVVTKYKVDRTQDELSTKYHCGIVFLDEVYALHMSCH 1237
DB 1214 HTGERPYTCTVCGKFFDRSSVVKH---SRHTGERPYKNECTGKGVOKSGLVXHMRT 1270
QY 1238 GDSGPFQCSICCHLCTDKYDFTTHIQRLHNRNAQVEXNKP 1279
DB 1271 TGEKPYGONCCOR-----SPSTHSASVHQ---RMCNTGRP 1303

RESULT 14
AREA GISFC
ID AREA GISFC STANDARD; PRT; 971 AA.
AC P79688;
CT 15-JUL-1998 (Rel. 36, Created)
CT 15-JUL-1998 (Rel. 36, Last sequence update)
CT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nitrogen regulatory protein area.
GN AREA.
OS Gibberella fujikuroi (Bakanae and foot rot disease fungus; Fusarium moniliforme);
CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OX Hypocreales; Hypocreales; Nectriaceae; Gibberella.
CK NCBI TaxID=5127;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ms67;
RX MEDLINE=599:69774; PubMed=10071216;

```

```
RA Tudzynski B., Homann V., Peng B., Marzluf G.A.:  
RT "Isolation, characterization and disruption of the area nitrogen  
RL regulatory gene of Gibberella fujikuroi.";  
RL Mol. Gen. Genet. 261:106-114(1999).  
CC -!- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN. POSITIVELY ACTING  
CC REGULATORY GENE OF NITROGEN METABOLITE REPRESSION (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- SIMILARITY: Contains 1 GATA-type zinc finger.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@sib-sib.ch).  
CC  
CC EMBL; Y11006; CAA71897.1; --  
DR HSSP; P17429; 4GAT.  
DR InterPro; IPR000679; Znf_GATA.  
DR Pfam; PF00320; GATA; 1.  
DR PRINTS; PR00619; GATAZNFINGER.  
DR SMART; SM00401; Znf_GATA; 1.  
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.  
DR PROSITE; PS00114; GATA_ZN_FINGER_2; 1.  
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;  
KW Nuclear protein; Nitrate assimilation.  
FT ZN FING 694 718 GATA-TYPE.  
SQ SEQUENCE 971 AA; 103580 MW; 887DD88214C7453 CRC64;  
  
Query Match 3.4%; Score 234.5; DB 1; Length 971;  
Best Local Similarity 21.0%; Pred. No. 8.7e-06;  
Matches 181; Conservative 98; Mismatches 313; Indels 269; Gaps 41;  
  
QY 344 GNSSTLEQHFLOTHPNKIKASLPSE--VAKSEKN-----SNKSIPLAQSDSGSL 394  
DB 186 GSRAFLGRPMQSSFGIAQLAKSENNAQADAVNLDGFISSNGSGSPINFAPEGD 227  
QY 395 -----GKQDKITVAGSDTP-----VGVSVPIKPLDSRRONGTEATSYVCKTCSFSC 443  
DB 226 MVDRRSGSSMASAIPKSRKEPSLQNFVQSPVQFAHOATQ-GSE-FYVVRNR-RKTSI 285  
QY 444 ESSSSLLKLEHYCKHCAVQSGGLNPLNDKLSRGVINGNDLAKSSEGETYKTKCKSS 503  
DB 286 DDRTRKRPANFSQVPAV-----NSTAQNLDLSELDHVSJLCPNQA 330  
QY 504 GAKKDFSSKGA-----EDNMTSYNQCFDFRYKSHGPDVIVGPI-LRHYQQ 552  
DB 331 GIPQQ---SNGSNVPEMIDTFEMENDSMVNNQFQ-QNFSFSPSTSP-MIPGHGFSQMY-- 383  
QY 553 LNIHKTIKHCFPCPGLCSPEKHGEITYPACRKNCSHCALLLLHLSPGAGSSRY 612  
DB 384 -HN-----SSVPSASMSNNNSD-----FSPPASAYPSNV 414  
QY 613 KHQCHQCSFTPDVDVLLPHVESVHESQASDVQKQEAHL-----QGSDGQQ-SVKESK 664  
DB 415 -----STPHVPVQEGFYFGSCQARTCPQGFQSTGSMLSQGFMYGNGNSGTYFSA 469  
QY 665 EHSTCKDFITQVEEISRHRYRAHSCYKCRQCSFTAAADTQSILLEHFNTHVCOEDITTA 724  
DB 470 PGTASE-----SMSAYSTAPSSFGHIDPSQVFQ-----EQAVTS- 504  
QY 725 NGEDGHAISTIKKEPKIDFRVNNLLTPDSKMEPVSESVVREKLEEKGLKEKWTES 784  
DB 505 -----PTIQPQNMFSGADSDDEDNAPADRVNSMQKD-----MSS 542  
QY 785 SSDLRNVTVRG-----ADLRGSPSYTCASLGILLTPVSGT----- 820  
DB 543 SLDESGANGDASLPGCFSTQAAAFPGGPRKQWNG-----GTTTDFVDNNGDNESNG 596  
QY 821 --QBQTKLRDSPNVEAH-----LARIYGLAVETKGF-----LQAPAG---GEKS 963  
DB 597 LERSQSQSFRRG-NLRQHPKLPENASTPVH-FGGQNGFEQLAQMSQSPAGDNGTWS 654
```

```
QY 864 G--ALPQVYPASGFNKSKDESQSLRRRRSG-----VFCANCLTTKTSLMWRNANGGY 915  
DB 655 GFSVAISRRSS-PPMSKGGSTTNLQANAGNGDNAPATCTTCNCTOTTFIWRNPEGOP 713  
QY 916 VQACGGYQKQKHSCTPRPLNLIKONNGECIIRPRTRKALNFALQALQKQCGSSNEE-Q 974  
DB 714 LQACGGFLXJHGVVAT-----SLKTDVKKRNRSGSTNVP 750  
QY 975 VQGSPLERREDEHLSHQREIPLPSLSKYEAQGSJTKSHSAO-----CPVLVSOTLQI 1028  
DB 751 VGSSTRSKZASTLSKRKNS-----TISMSTATANSTKPNSSNPTPRVTPPATSQ----- 802  
QY 1029 HKRWQPLHDKKQPEST-----GDPGNSSSVSEKSGSSRGSPIEKYMR 1073  
DB 803 ----PPSSKQVDSPVSTTSQANTAGSTPNSHFHGGFGPSSGAVGGKGV-----VPI----- 849  
QY 1074 PAKHPNVSP-FGSPIEKYQYP 1093  
DB 850 AAAPKTSPPGAGSSMSMQRP 970  
  
RESULT 15  
TKAR_VQJUSE  
ID TKAR_VQJUSE STANDARD; PRT; 517 AA.  
AC Q03267; Q64044; Q64045; Q64051;  
DT 01-OCT-1993 (Ref. 27, Created)  
DT 15-DEC-1998 (Ref. 37, Last sequence update)  
DT 15-SEP-2003 (Ref. 42, Last annotation update)  
DE DNA-binding protein Ikaros (lymphoid transcription factor Lymph-1).  
GN ZNFN:AL OR IKAROS OR LYF.  
OS Mus musculus (Mouse).  
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI_TaxID:10090;  
PN 11 TaxID:  
RP SEQUENCE FROM N.A. (ISOFORM V).  
RC TSSJB=Embryo;  
RX MEDLINE=91368267; PubMed=1439790;  
RA Georgopoulos K., Moore D.D., Deifler B.;  
RT "Ikaros, an early lymphoid-specific transcription factor and a  
RT putative mediator for T cell commitment";  
RL Science 258:808-812(1992).  
RN 12)  
FP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ALTERNATIVE SPLICING.  
RA MEDLINE=95021239; PubMed=7935426;  
RA Hart K., Ernst P., Lo K., Kim G.S., Turck C., Smale S.T.;  
RT "The lymphoid transcription factor Lymph-1 is encoded by specific,  
RT alternatively spliced mRNAs derived from the Ikaros gene";  
RL Mol. Cell. Biol. 14:7111-7123(1994).  
CC -!- FUNCTION: BINDS AND ACTIVATES THE ENHANCER (DELTA-A ELEMENT) OF  
CC THE CD3-DELTA GENE. FUNCTIONS IN THE SPECIFICATION AND THE  
CC MATURATION OF THE T LYMPHOCYTE. ALSO INTERACTS WITH A CRITICAL  
CC CONTROL ELEMENT IN THE TCT (TERMINAL DEOXYNUCLEOTIDYLTRANSFERASE)  
CC PROMOTER AS WELL AS WITH THE PROMOTERS FOR OTHER GENES EXPRESSED  
CC DURING EARLY STAGES OF B AND T CELL DEVELOPMENT.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoform=s6;  
CC Name=VI;  
CC IsoId=Q03267-1; Sequence=Displayed;  
CC Name=I;  
CC IsoId=Q03267-2; Sequence=VSP_006855;  
CC Name=II;  
CC IsoId=Q03267-3; Sequence=VSP_006853, VSP_006855;  
CC Name=IV;  
CC IsoId=Q03267-4; Sequence=VSP_006856;  
CC Name=V;  
CC IsoId=Q03267-5; Sequence=VSP_006853, VSP_006856;  
CC Name=VI;  
CC IsoId=Q03267-6; Sequence=VSP_006854;  
CC -!- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN T-CELLS AND THEIR  
CC PROGENITORS, AND ALSO IN B-CELLS.
```

```
CC -!- SIMILARITY: BELONGS TO THE IKAROS FAMILY OF ZINC-FINGER PROTEINS.
CC -!- SIMILARITY: Contains 6 C2H2-type zinc fingers.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L03547; AAB66193.1;
DR EMBL: S74517; AAB32248.2; ALT_SEQ.
DR EMBL: S74518; AAB32249.2; -.
DR EMBL: S74708; AAB32250.2; -.
DR HSSP: P15822; 1BBO.
DR TRANSFAC: T01470; -.
DR MGD: MGI:1342540; Znf141.
DR GO: GO:0003677; F:DNA binding activity; IDA.
DR GO: GO:0003097; P:hemopoiesis; IMP.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IMP.
DR InterPro: IPR007087; Znf C2H2.
DR Pfam: PF00096; zf-C2H2; 5.
DR ProDom: PD000003; Znf C2H2; 2.
DR SMART: SM00355; ZNF_C2H2; 6.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 3.
KW Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW DNA-binding; Nuclear protein; Repeat; Alternative splicing.
FT ZN_RING 117 139
FT C2H2-TYPE 1.
FT ZN_RING 144 166
FT C2H2-TYPE 2.
FT ZN_RING 172 194
FT C2H2-TYPE 3.
FT ZN_RING 200 223
FT C2H2-TYPE 4.
FT ZN_RING 457 479
FT C2H2-TYPE 5.
FT ZN_RING 488 512
FT C2H2-TYPE 6.
FT VARSPPLIC 53 53
FT M -> VAYGAGGDRDFHALISDRGM (in isoform I)
FT and isoform IV).
FT /FTID=VSP 006853.
FT Missing (in isoform VI).
FT /FTID=VSP 006854.
FT Missing (in isoform I and isoform II).
FT /FTID=VSP 006855.
FT Missing (in isoform III and isoform IV).
FT /FTID=VSP 006856.
FT VC -> MY (IN REF. 2).
FT CONFLICT 234 235
FT CONFLICT 480 482 MISSING (IN REF. 2).
FT SEQUENCE 517 AA; 57336 MW; 1052B8E76AF24287 CRC64;

Query Match
Best Local Similarity 23.6%; Pred. No. 3.9e-06;
Matches 127; Conservative 63; Mismatches 204; Indels 144; Gaps 25;

QY 814 LTFVSGTQQTTLRD-SNNVEAAHLARIYGLAVETKGFLOGAPA---GGEKSGALPQQ 869
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
37 LSTTGAQGNKSDRGMSNVK-----VETQSDENGRACENNGEACEDLRM 84
QY 870 YPASGENKS---KDESQSLRRRG-----SQVFC--ANCLTKTSLWRKNAN 912
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
85 LDASGEKXNGSHRDQGSALSGVGGIRLPNGKIKCDICGIVCIGPNVLWYHK---RSHT 141
QY 913 GGVVNCAGJ-----YQKLHSTPRPLNIKKNGEQIIRRTKRLNPEALQAEQ 962
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
142 RPFQCNQCGASFTQXGNLLRHILKHSGEK---FKCHLQYACRR-----DALTCHL 191
QY 963 LNKQQRGSNEEQVNGSPLERS--EDHLTESHQ--REIPJPSL-----SKYEAQSSL 201C
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
192 RTHSVGKPHKCYGGRSYKQRSLEBHKRCHNYLESMLPGVCFVKEETNNHNEAEDL 251
QY 1011 TKSHSAQQPVLYSQTLDIHKWOPHLIQIKSQESTGDPGNSSSVSEKSGSEROSPTEK 207C
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
252 CKIGAERSLVLDRLASNVARKSSM-----PQKFLGDKCLSDXPYDANYEKEDMTSH 305
QY 1071 YMRPAKHP--NY-----SPFGS-----PIEKYQY-----PLFGLPFVHNDQSEA 1138
```

Search completed: October 29, 2003, 11:30:53  
Job time : 44 secs

```
Db 306 VNDQAINNA:NYLGAESURPLVQTPPGSSSEYVFISSMYQLHKPPSDGPPSRNSACDAV 365
QY 1109 DWLREWSKYKLSVPCNPHYLSHVPGLP-NPCQNYVP-----YPTFNLPEHFS 1154
Db 366 DNLLLSKAK-SVS-----SEREASPSNSQSDTDIESNAEQRSGLIYLTNH-NP- 416
QY 1155 AVGSNDIPLDLA:KHSRPGPTAGASKEKTKAPFNKNEGPNVYVTEKYDRSTQDELS 214
Db 417 -----ARNGLALKEECRAVEVLRAASENSQDAFRVVSSTGEOLK 455
QY 1215 T-KCVHCGSIVELEFNVALHMSCHDSQ---PPOCS:QCHLCTCKYCPFTTH:QRLHR 1268
Db 456 VYKCEHCPVLFJLHVMY-TIHMGCHGCHGFRPFECNCGYSGQRVYFSSK:TRGEER 513
```

1	5257.5	76.7	1092	11	Q8K1J0
2	282	13.5	229	11	Q8B262
3	983	4.1	1852	11	Q8CH66
4	271	4.0	1186	11	Q08961
5	267.5	3.9	1224	4	Q9NZ13
6	265.5	3.9	1311	11	Q8C102
7	265	3.9	1271	11	Q8C1Q1
8	264.5	3.9	1311	11	Q8VD56
9	263.5	3.8	1669	11	Q22535
10	262.5	3.8	589	13	Q8UVK1
11	262	3.8	1167	11	Q9ESD2
12	261.5	3.8	1669	11	Q925V4
13	260	3.8	537	13	Q93581
14	258	3.6	507	13	Q918U5
15	256	3.7	1311	4	Q96K83
16	256	3.7	1412	4	Q96JM2
					Q8K1J0 mus musculus
					Q8B262 mus musculus
					Q8CH66 mus musculus
					Q08961 rattus norv
					Q9NZ13 homo sapien
					Q8C102 mus musculu
					Q8C1Q1 mus musculu
					Q8VD56 mus musculu
					Q22535 mus musculu
					Q8UVK1 petromyzon
					Q9ESD2 mus musculu
					Q925V4 mus musculu
					Q93581 brachydanio
					Q918U5 petromyzon
					Q96K83 homo sapien
					Q96JM2 homo sapien

1 VVR3KNPP\*.BNVASEGEGOT\*.EPTATESKVSCKNKFELSAJCKXSENTCCSNVAE\*.NSKEEF. 60

```
QY 61 SLHVQDPSSSSKKDLKSAVLSEKAGFNYESPSKGNFSPFFPHDEVTDNRNLAFSPFAAG 120
DB 61 STHGQEPSSSGKKDLQISGLSEKAGFNYESPSKGSLSVFFPHDEVTDNRNLAFSPFAAG 120
QY 121 VCEPLKSPORABADDPQMACTPSSGDSLETXEDQKMSPKATEETGQAQSGQANCGGLSPV 180
DB 121 VCEPLKSPORABADDPQMACTPSSGDSLETXEDQKMSPKATEETGPVQSGQANCGGLSPV 180
QY 181 SVASKNPQVPSDGGVRLNKSITDLLVNPNPPAPLSPRELQPKNIQGYGYGNPDTCLI 240
DB 181 SVASKNPQVPSDGGVRLSKPGDLLVNPNPPAPLSPRELQPKNIQGYGYGNPDTCLI 240
QY 241 XHFRKYHLGLHNRTRQDAELDSKILALAHNMVQFSKDFQKVRNSVPSGVLLQD:NSSRPV 300
DB 241 XHFRKYHLGLHNRTRQDAELDSKILALAHNMVQFSKDFQKVRNSVLSGVLLQD:SSSRFA 300
QY 301 LINGTYDVQVTSGGTFIGRKTTPCOGNTKYPRCKFCNFTYNGNSSTELBOHFLQTHPN 360
DB 301 LINGTYDVQVTSGGTFIGRKTTPCOGNTKYPRCKFCNFTYNGNSSTELBOHFLQTHPN 360
QY 361 KIKASLSPSEVAKPSEKSNKSIIPALQSDSDGJLCKWQOKTVKAGDDTPVGYGVPIKPL 420
DB 361 KIKVSLSPSEGVKPEKSNKSIIPALRASDGVCKWQOKTVKAGDDTPVGYGVPIKPL 420
QY 421 DSSRQNGTEATSYWCKFCSPSESSSSKULEHYGKHGAVQSGGLNPELNDKLSKGV 480
DB 421 DSSRCV----- 426
QY 481 INQNDLAKSSBGETWTKDKSSSGAKKXDFSSKGAEDNMVSYNCCPCDFRYSKSHQPDV 540
DB 427 ----- 426
QY 541 YVGPLRLHYQOLHNIHKTIKHCPCPCRGCLSPCKHLIGE.TYFPACKRKNCSHCALJLL 600
DB 427 ----- 426
QY 601 HUSPGAASSRKYKHQCHQCSFTTPDVVLLFHYESVHESQASDVQKQENHLOGSGGQSV 660
DB 427 -----DVLLFHYETVHESQASDVQKQENHHLGSGGQAV 460
QY 661 KESKHSCTKCDPITQVEEISRHVRRRAHSCYKQCQCSFTAADTOSLLEHNTVHCQSD 720
DB 461 RSKSHSCTKCDPITQVEEISRHVRRRAHSCYKQCQCSFTAADTOSLLEHNTVHCQSD 520
QY 721 ITTANGEBGHAISTIKPEPKIDFRVNYLLTPDSKMGEPVSESVYKREKLEKQGLKEV 780
DB 521 ITTANGEBGHAIPTIKEPKIDLVYSLNPNDSXMGETVPESIVKREKLDQKEGJAKXI 580
QY 781 WTSSSSDRLNVTMRGADILRGSPSYTOASLGLLTPVSGTOETKTLRDSPNVEAAHLAR 840
DB 581 WTSSSTDDLRGVAMRGADILRGSPSYTOASLGLVLPVSSSQEOTKTLRDSPNVEAAHLAR 640
QY 841 PTYGLAVETKGLQAPAGKSGALPQOYPASGENKSKDESQSLL-----RRR 899
DB 641 PWYGLAVDTKGLQAPAGSEKASLTQOYPASGSKTKDESQSLLRAVTSQTPNRRR 700
QY 890 RSGGVPCANCLTTKTSLRMKNANGVYVNCAGLYOKLHSTPRPLNIKQNGEIOIIRRT 949
DB 701 RSGGVPCANCLTTKTSLRMKNANGVYVNCAGLYOKLHSTPRPLNIKQNGEIOIIRRT 760
QY 950 RRLNPEALQABLNRKQNGSNEOVNGSPLERRSECHLTESHOREIPLPSLSKYEACGS 1009
DB 761 RRLNPEALQABLNRKQNGSNEOVNGSPLERRSECHLSHREIPLPSLSKYEACGS 820
QY 1010 LTKSHSAQQVPLVSQALDIHKRMQPHIQIKSPQESTGDPGNSSSVSEKSSSERGSPIS 1069
DB 821 LTKSHSAQQVPLVSQALDIHKRMQPHIQIKSPQESTGDPGNSSSVSGKSSSERGSPIS 880
QY 1070 KYMRPAKHPNYPSPGSPTEKYQYPLFGLPFFVHNDQSGEADMLRFMSKYKLSVPQKPHYLS 1129
DB 881 KYMRPAKHPNYPSPGSPTEKYQYPLFGLPFFVHNDQSGEADMLRFMSKYKLSVPQKPHYLS 940
```

```
QY 1130 HYPGLPNPCQNVVPTFNLPHPHSAVGSNDIPLDLAIKHSRPOPTANGASKETKAPP 1189
DB 941 HYPGLPNPCQNVVPTFNLPHPHSAVGSNDIPLDLAIKHSRPOPTANGASKETKAPP 1000
QY 1190 NYKNEGPPNVVYKTEKVDNSTQDELSTKCVHCGIVFJDEVYALHNSCHGDSGPPFCCSICQ 1249
DB 1001 TVKNEGPPNVVYKTEKVDNSTQDELSTKCVHCGIVFJDEVYALHNSCHGDSGPPFCCSICQ 1060
QY 1250 HCTDKYDFTTHIQGLHNNNAQVEKNGKPKKE 1281
DB 1061 HCTDKYDFTTHIQGLHNNNAQVEKNGKPKKE 1092
RESULT 2
Q8Z52 PRELIMINARY; PRT; 229 AA.
ID Q8Z52 PRELIMINARY; PRT; 229 AA.
AC Q8Z52;
DT 01-VAR-2003 (TRENDELrel. 23, Created)
DT 01-VAR-2003 (TRENDELrel. 23, Last sequence update)
DT 01-VAR-2003 (TRENDELrel. 23, Last annotation update)
DE Trichorhinophthalangeal syndrome 1.
CS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN 1.
PP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Hole;
RX MEDLINE=23354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team.
RT "Analysis of the mouse transcriptome based on functional annotation of
RL Nature 420:563-573(2002).";
DR EMBL; AK036590; BAC29495.1;
SQ SEQUENCE 229 AA; 25290 MW; 03283E1B7AABA039 CRC64;
Query Watch: 13.5%; Score 923; DB 11; Length 229;
Best Local Similarity 85.6%; Pred. No. 2,3e-54;
Matches 173; Conservative 17; Mismatches 12; Indels 0; Gaps 0;
QY 695 RRAHSCYKQCQCSFTAADTOSLLEHNTVHCQSDITTANGEBGHAISTIKESPKIDFR 745
DB 20 RRAHSCYKQCQCSFTAADTOSLLEHNTVHCQSDITTANGEBGHAISTIKESPKIDK 79
QY 746 VYNLLTPDSKMGEPVSESVYKREKLEKQGLKEVYKWTSSSDRLNVTMRGADILRGSPS 805
DB 90 VYSLNPNDSKMGETVPESIVKREKLDQKEGJAKXIWTSSSTDDLRGVAMRGADILRGSPS 139
QY 506 YTOASLGLLTPVSGTOETKTLRDSPNVEAAHLARPIYGLAVETKGLQAPAGKSGA 865
DB 140 YTOASLGLVLPVSSSQEOTKTLRDSPNVEAAHLARPIYGLAVDTKGLQAPAGSEKAS 199
QY 666 LPOQYPASGENKSKDESQSLLR 887
DB 200 LPOQYPASGENKSKDESQSLLR 221
RESULT 3
Q8CHH6 PRELIMINARY; PRT; 1852 AA.
ID Q8CHH6 PRELIMINARY; PRT; 1852 AA.
AC Q8CHH6;
DT 01-VAR-2003 (TRENDELrel. 23, Created)
DT 01-VAR-2003 (TRENDELrel. 23, Last sequence update)
DT 01-VAR-2003 (TRENDELrel. 23, Last annotation update)
DE KIAA0236 protein (Fragment).
GX KIAA0236.
CS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN 1.
PP SEQUENCE FROM N.A.
```

RC TISSUE=Brain;  
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Hara Y., Nagase T.,  
RA Ohara O., Koga H.;  
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:  
RT I. The complete nucleotide sequences of 100 mouse KIAA-homologous  
RT cDNAs identified by screening of terminal sequences of cDNA clones  
RT randomly sampled from size-fractionated libraries";  
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB093219; BAC41402.1; -;  
FT NON\_TER 1  
SQ SEQUENCE 1852 AA; 207196 MW; 4AED7B66EE130019 CRC64;  
Query Match 4.1%; Score 282; DB 11; Length 1852;  
Best Local Similarity 18.2%; Pred. No. 8.3e-10;  
Matches 257; Conservative 153; Mismatches 447; Indels 552; Gaps 63;  
QY 156 YSPKATEETGCA-----QSGAQCGLSPSVASKNPQV 189  
DB 93 LSPGAPETSGVLVKKVVEYFECERCEQSFABETLLSVHQCTETHIQAVQDLSPPCSVEL 152  
QY 190 PSDGGVRLNKSXTDLLVNDNPPAPLSPPELQDFKCNICGYYGNDPTDLIKHFKYHJG 249  
DB 153 P-PSNALRGPLQDPSLPPLP-----CPVCRQEFV--QPQALKSHFK----- 193  
QY 250 LHNRTQDAELDSKILALHNWQFQSHK--DFQKVRVSVPFGVLQDINSRPV----- 300  
DB 194 IHRVT-----PNMFSPESCVCFAEDRKLQNLHRLQTHKAVFPVPCFRCGS 240  
QY 301 LLLNGTYDVQVTSGGTFIGIGRTYPCQGNTKV--PRCKFCNFTYMGNSSTELQHFQCTHP 359  
DB 241 LLFGS-----QQGMELHQAHYFPFCHSCFP--MGSNVKLFRQH--QRSHG 282  
QY 360 NKIKA-----SLPSSEV-----AKPSEKSNKSPALQSSDGLGRQDKIT 402  
DB 283 ASARGELSAVQGLPSQELLPAKLPPGHRPESSEAST--PLPGQESAEEED--APEEESVT 340  
QY 403 VKAGDDTPGVSYVPFKPLDSSRQNGTEAT--SYVKCKFC-----S 440  
DB 341 QKDSQKV--MDKSGGAQQLGHEGVSGTESLFKTHMCPCKRCFKKTHLVEHLH--HFPDPS 399  
QY 441 FSCES-----SSSLKLEHYGQHGAVQSGGGLNPELNDK--SRGSVINONLAKSSEGSTV 495  
DB 400 LQCPNCKQKFTSKSLKTH-----LRLGELKAHRCPLCHYSVAVERNALNRYH 447  
QY 496 TKTDSSGAKKDKFSSKGAEDNMVTSYNCQPC--DPRYS-----KSHGPDVTV----- 542  
DB 448 ASMHEDIS-----NFYS-----DTYACPVCREEFLSLQALKHLSHT--AAAAAEPFLP 494  
QY 543 -----VGP-----LRLHYQDLHN1-----HK----- 558  
DB 495 LHCFOEGCTYVAPDRKAFJLKLKBINGVRAVECRHHSCFPLPATABAEAHKSHYATHC 554  
QY 559 -----CTIKH-----CPFCPRGLCSP-----EKHLGEI--TYPPACRKSX 591  
DB 555 PHCDACSNKHLFRKHKKQGHGPGSEBELRCTFPFPATFNPAVQDHVGXMAHEKTHQCE 614  
QY 592 C-----SHCALLLHLLSPGAGSRVNHQCHQCSFTTPDVTVL----- 629  
DB 615 CNFATAHKRVJLRHM-----LLH7GKPHKCELCDFTCRDVSYLSKMLTHSNKTDYMCYEC 671  
QY 630 -----LFHYESVH--ESQASDVQKEAN-----HLOGSDGQSVK-----ESKHSCTKCFITQ 676  
DB 672 GYVTKWKLVSVMRKHAGDLRYQCQCYSYRCHRADQLSSHLRHQGSKSLNCEVCAFAK 731  
QY 677 VEEIISRHRYRAH-----SCYKCRCSFTAADTOSLLEHN-----TVHCQCDIT 723  
DB 732 RXYELQKMASQHHPTGTPAPLFCRCVCSYQSRHKQALLSHENCKTHLREHFCA--CDYRT 791  
QY 724 AN-----GEE-----DGHASTIKSEPKIDFR 745  
DB 792 FSNITLFFHKRVKHVGYMPGDQVQWQFNCASQELGAPQCCLAPSDSGPSSQLSAQPEREDR 851  
QY 746 VYNLL---TPDSKMGZPVSESVVKREKLEEKQGLKEKWTESSDOLRNVTWRGADILRG 802

DB 852 EHEIVANSNMQALPETNEEAFKRC-----DGIE-----APQSDQVD----- 890  
QY 803 SPSTYTCASLGULTFVSGTCEQTKLRD--SENVEAAFLARFIVGLAVETKGFLOGAPAG 860  
DB 891 SPISLGEVWG-----SCTLHLEALVLELEPETEPLEELTETATVEFRPLDPSGLGT 944  
QY 861 EKSGALPQQVPASGENKSKDESQSLRRRRRSGVFCANCLTTKTSLMRKXNANGVYVCNAC 920  
DB 945 ERPGGLEF--PALSSPDS--IETPALVAESE----- 97;  
QY 921 GLYQKLNSTPRPLNI--KQNGEQ--IARRTRKSLN--PEALCAQOLNKQCSNEEQVNSGP 979  
DB 972 PVYVKLASEPPRNPPLISEBAPTTFKAALTAETVPJPFPESESLDKAMRRQDKEQAEALV 1031  
QY 960 LERRSE-----DHJTESHQREIPLPSLSKYBAQGSJTKSHSAQQPVLYSQTU 1026  
DB 1032 LEGRVQWVVIQEGRAFRCFPCFPFTRREKALTLSKSGCG-----RREPLLCECG 1084  
QY 1027 DIHKRQPL--H7QIKSPOESTGDPNSSSVSEKSGSSRGSPDKYMRPAKHP----- 1078  
DB 1085 ASFKQRLGLSTHMKKCPV-----LKKNKALPKEVSPFTLHPQ--PDNQ 1127  
QY 1079 -----NYSPPGSPIEKYQYPLFGLFVHNDFOSEADLRFNSKYKLSVPGN--PHYLSHV 1131  
DB 1129 ASQDAESRKPPPLP-----SKVELLPKDPSPSLPGG 1159  
QY 1132 PGLPNPCQNVVYPTFNLPHPFSAVSGNDKIPDLAIKHSRPGFTANGASKKXTKAPPNV 1191  
DB 1160 FGVEELPTSDFT--SP-----ENSLP-----TGTEK----- 1188  
QY 1192 KNEGPLNVWVTEKVDNRSTQDELSTKVCVHCIVFDEVMYALHVSCHGDSGPPQCS--CQHL 1251  
DB 1169 -----FHFEQKPHCSSCTFL 1204  
QY 1252 CTDKYDFTTHIQRLHNNNAQVEKNGKPK 1280  
DB 1205 CSRSLSSITSHVTEGCRGGRGQKRGKGRPQ 1233  
RESULT 4  
Q0836;  
ID Q0836; PRELIMINARY; PRT; 186 AA.  
AC Q0836-1;  
DT 01-JUL-1997 (TREMBLrel; 34, Created)  
DT 01-JUL-1997 (TREMBLrel; 34, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel; 23, Last annotation update)  
DE Roaz.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1] SEQUENCE FROM N.A.  
RC STRAIN=Sprague-Dawley;  
RA Tsai R.Y.H., Reed K.R.;  
RT "Cloning and Functional Characterization of Roaz, a Zinc Finger Protein  
RT that Interacts with C/EB-1 to Regulate Gene Expression: Implications  
RT for Olfactory Neuronal Development";  
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; J92564; AA558646.1; -;  
DR HSSZ; PC8047; JSP1.  
DR TRANSFAC; T02815; -;  
DR InterPro; IPR007387; Znf\_C2H2.  
DR Pfam; PF00036; zf-C2H2; 28.  
DR ProDom; PDC00003; Znf\_C2H2\_2.  
DR SMART; SM00355; Znf\_C2H2; 28.  
DR PROSITE; PS00028; ZINC FINGER C2H2\_1; 27.  
DR PROSITE; PS00157; ZINC FINGER C2H2\_2; 21.  
KW Metal-binding; Zinc; Zinc-finger.  
SQ SEQUENCE 1186 AA; 133479 MW; 562D6C779BC9BFCA CRC64;  
Query March 4.0%; Score 271; DB 11; Length 1186;





QY 255 ---ROAELDSKILALHNMVQFSKQKQV- 283  
Db 210 DTFSQTELEKHLVTRH--POLSEKADLQC/HCPVFDVENTLLAHIIHQAANQKHKCPM 267  
QY 284 ---RSVFSGLQDINSR-----PVL-----LNGYD--VQTSST 315  
Db 268 CPEQSSVEGVYCHLDSHRQPDSSNHSVSPDVLGVSMSASATPDSSASVERGSTPDST 327  
QY 316 ---FIGIGRKTDPCCGNK-YPRCKFCNFTYMGNSSTELQHFLOTHPNK/KA----- 364  
Db 328 LKPLGQKMRDDGGGATVYVSCYCKRDF-NSLAVIEHLKT/HADKQCGH/COIC 386  
QY 365 ---SLPS-----SEVAKPEKSNKSPALQSSD-----SGDLGKQCDXITVX-A 405  
Db 387 LDSMPTLYNLNHRVKLHK--NHAYPVQVQGNISAFHCNYPCEMFADINSLOEIRVSHC 444  
QY 406 GDDTFVGVSPKIP/LDSSRQNGTEATSYWCKFCSCFSCSSSLKLLHRYKQKCAVOSG 465  
Db 445 GPNA-----NPSDGNNAFFCQCSMGFTJESS--LTHHIOAHCSVCSA 486  
QY 466 GLN-----PELNDKLSRGSVT-----NONDL-----AKSSEGETNY 496  
Db 487 KLESPVQVQTSFMEVSCYCTNSPIFGSILKTKHIKENHKNIPLAHSKSKAEQSPV 546  
QY 497 KTDKSSGAKKXDFSKGAEDNMVTS--YNQFCDFRYSKSGHPDVIVVGLJRHYYOOLH 554  
Db 547 SSDVEVSPKQRLS---ASANSISNGEYPCNQCDLKFNFES-----FQHLKLIH 594  
QY 555 NIHKCTIKHCP/CPGLCSPEKHLGEITVPPACKSN--CSHC-----ALLL 600  
Db 595 LELLURKQACPOCKEEDFQESLLOH/LVHYMTTTHVYVCSCKQFSSVDLQKHLLDM 654  
QY 601 HLPGAAGSRVHVOHQCSFTTPD/VDLLPHYESVHESQAS-----DVQEAHNL 651  
Db 655 H-----TFVLXHTLCOEVPFDSKVSIOVHLAVKSNKXKRYCTACNWPFRKEA--- 703  
QY 652 QGSDQQQSVKES-----KEHSCTKCDFTQVVEBISRYRAHSCYKRCOSFTAAQTQ 705  
Db 704 ---DLQVHVXHLGNPAKAHAKICFCGETFSTEVELQCHITTHSKYKCKFCSKAFHAI 760  
QY 706 SLLEHFNTHCOEQDIT/TANGEDG-HAIST/KEPKIDFRVYNLLTPDSXNGEPYSV 764  
Db 761 LLEKHLREKHC-VFAATENGANGVPPMATKAEP-ADLOQM/LKNPEA---PNSH-- 812  
QY 765 VKREKLEKDG/KEKVTWTESSDDU--RNVTRGADILRGSPYTCAS/LLLTPVSGTQE 823  
Db 813 -----EASEDDVDASEPMYGCDC--GAATVYEVL-----LQ 842  
QY 824 7KTLRSPNVEAAHLARP/YGLAVET/KGFLQAPAGGKSGALPOQYPASGENSKSDESQ 883  
Db 843 NHRLRD-----HNIRP-----GEDDGRKKAETIKSHKNCVSR 877  
QY 884 SLRRRGSGVFCANCLITKTLRKNANGVYVQACGLYKHLSTPRPLN/IKONGEQ 943  
Db 878 TP-----PSENGLREHLTHRGPAK-HYMCPICG-----EFPFLTLTEH----- 917  
QY 944 IIRRTKRLNPEALQAEGLNQSGSDEQVNGSFLERRSEDHUTESHQREIP/PSLSK 1003  
Db 918 -----KVTHSKSLDTGTRICKMPLQ--SEEFIEHCQMH---PDL-- 953  
QY 1004 YEAQSGLT--KHSAAQPVLVSGTLDIHK--RQPH/HTQIKSPQSTGDPGNSSSVSEKG 1059  
Db 954 ---RNSLTGFRVCVCMQTV/STLELKIHTGTFHMQL-----AGSSAAS--- 993  
QY 1060 GSSRGSPIEKYRPAKHENYSPGSPIEKYQYPLGFLPVHNDFQSEADNLRPFMSKYL 1119  
Db 994 -----SPNGGGLQK-----LYKCALCLKEFRSKQDLV-----KJ 1022  
QY 1120 SVPGNPHYL-----SHVPLG--PNPQCNVYVPTFNLPHPFSVAVGSDNDIPLOLA 1167  
Db 1023 DVNGLPYGLCAGMARSANGVGLAPPAPDR--PCAGLR/PECSVVFESADLESYK 1080

QY 1168 IZHSRPGPTANGASKKTKAP-PNVKNEGLNVKTEKVDORSTQ-----DELSTYC 1217  
Db 1281 VZHRDLTPTSGPRKQTQTSVPBRKXYQCIKQMTFENERETQIHVANKYFEGINHEC 1340  
QY 1218 VHCGLVFLDEWVALHM---SCHGDSGPFQCS/C-----QEL----- 1251  
Db 1141 KLCNQMTFSAKLCHLIEHSEFGXGTFCPCVCTVFVQANKLQHF/FAVHGCSDKIYD 1200  
QY 1252 ---CTKDYFTTHIQRGJHRNNAQ 1272  
Db 1201 CSCCPKQFFQTELOQNE/TMSQHAQ 1224  
RESULT 6  
Q8C1Q2 PRELIMINARY: PRT: 1311 AA.  
AC Q8C1Q2  
D 01-MAR-2003 (TIEMBLure: 23, Created)  
D 01-MAR-2003 (TIEMBLure: 23, Last sequence update)  
D 01-MAR-2003 (TIEMBLure: 23, Last annotation update)  
DE Ecotropic viral integration site 3.  
GN EVI3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RF SEQUENCE FROM N.A.  
RC STRAIN=BALB/C;  
RA Warming S.; Jiu P.; Suzuki T.; Akagi K.; Lindtner S.; Pavlakis G.N.;  
RA Jenkins N.A.; Copeland N.G.;  
RT "Evi3, a common retroviral integration site in murine B-cell lymphoma,  
RT encodes an BBFAZ-related kruppel-like zinc finger protein.";  
RL Blood 0:0-0(2002).  
DE EMBL: AY147426; AAK39819.1;  
SQ SEQUENCE 1311 AA; 147623 MW; 53810EC6345C3D37 CRC64;

## Query Match:

3.91; Score 265.5; DB 11; Length 1311;

Best Local Similarity 19.24; Pred. No. 6.3e-09;

Matches 302; Conservative 177; Mismatches 516; Indels 575; Gaps 82;

QY : MVKVKQPP/ENVASEGEGQLEPIGTESKVGKKEFSADQKSENTDOS-DAA-----EL 54  
Db : MRRKQAKPRL-KDPYKLEKIEDGEAVDCKRPEDGEELEEDAVHSCSLQVFESL 59  
QY 55 NHKEERSLH-----VQC-----PSSSKKDLKSAVLSEKAGFVYSPSKGNP 98  
Db 60 SDTEHKLHCCQ/TCGVVDEDDPSCSNPASPSSSKDT/SPSHGEGCCFGE--EGG--P 114  
QY 99 SFPHD-EVTDNRNLAFFPAAGVCEPLKSPORAEADPDQMACTPSCDSELTKEQKMS 157  
Db 115 GJFVPCQFCDSFSRLSY-----LKHHEGSHS-DKLPPFKCTYCSRLFKHKRGRH 164  
QY 158 PRATE/GOAQSCANCCG-----LSPSVASKNP---QVPSGGVRLNKSXTDL 205  
Db 165 IKC--HTGDKKXVHCSCDAAFSSRSLXHLKTH/TSNKPYKCAVCRGFLSSSLHGMQ 222  
QY 206 VNC-NPDPAP/SPLEODF-----KNCIGYGYGNDPTDLIKHFKYH----- 247  
Db 223 VHERKXGSGSGMEDCKMKCTCKSCQCEGF--DFPEDLQKHIAECHPECSNEDRAA 280  
QY 249 ---LGLHNR/RODAE/LSKILALHNVQ-----FSSKSKQKQKVRNRF 287  
Db 281 LQCMYCHELFVEETS/LMNHIEQVGEKKNKSCISCESELTVEELYSHMDSHQOPESC-- 318  
QY 289 SGV/QDINSRPV/LN-GTYCVVTSSTGTFICI-----GRK-----TPD 325  
Db 339 -----NHNSPSLVTGVTSVSTTPDKNLSVCSSTWVEAAPPIPKSRGRKRAAQ/SD 392  
QY 326 CQGNKTK-----YPRCKFCNFTYMGNSSTE/EQHFLQTH---PJK/KASLPSSEVAKPSEK 377  
Db 393 MTGFSKQAKVTYVSCIYCN-KQLFSSLAVLQJHLKTNH/LDKPQAHICQYCLEVL-PSLY 450

```
QY 378 NSNKSIPAL-QSDPSG-----DLGKWDKITYKAGDTP-----410
D 451 NLNHLKQVHEADPGLIVSAMPALVQCNCFCSEWVNDLNTLOBHIRCSHGAFNPAKOS 510
QY 411 -----VQSVPIKPDSSRON-----GT-----EATSYKCKFC 433
D 511 NAFFCPHCYMGFLT-----DSSLSEHTRQVHCDLGSRGFSVGLGTPKEPVVEVYSCYC 565
QY 440 SFCSESSSLKLLBHYGKHGAVOGGLNPNLNDKLSRG-----SVINQNDL-----456
D 566 TNSPFFNSVLKLNKHKHKNIPALUNYTHNKKGRALSPLSPVAIEQTLLKXWGTVG 624
QY 487 ---AKSSEGETMTKTDKSSSGAKKDESS-----KGABDNMVTSYNQCDFRYSKSHGP 538
D 625 GGPARRAS-GEVIC---NOCCAKVTSLDSFQHLKTHLDTVLPLKJTCPCQCNKEPFNCES- 678
QY 539 DVIUVGPIIRHYQOLHNHKTCKHCPCPCPLGCLSP---EKHLGEI-TYFACRKSNCVSH 594
D 679 -----LLKREV-TIHFMTITTYICESCDQKFTSVQDLQKHLJCMHTFVF---FRCFL 726
QY 595 C-----ALLLJLSPGAAGSRVKKHCHOCSTTPDQVLLPHYBSVHESQASDYKQE 647
D 727 CQEVFDSKVSQQLHLAVKSHKVV-YECTSNWDFRNETLQJH-----VXH- 773
QY 648 ANHLGGSDGQOSVKSKEHSCTKC--DPTQVEEE--ISRHYRAHSCYKRCQCSFTAAD 703
D 774 -NHLE-----NCGKVHKCIFCGESFGTEVE-QCHITTHSKX-----YNCRPCKAFHA 820
QY 704 TOSLLEHNTVHCQEQQITANGEDGHAISTIKKEPKIDFRVNNLTPSKNGEPUSES 763
D 821 VILLEKHILREKHC-VFEKTPNCTNGTNGASEQVQKEAELO-----TLTN-----864
QY 764 VVREKLEBEKGLKEKVMYTESSSDLDNRVTRGADILRGSPSYTQASLGLLTPVSGTQE 823
D 865 --SOESHNSHDGSEEDV---DSSEPM-----YGCJIC--GRAVITETL-----LQ 902
QY 824 TKYLRDSDNVEAHLARP-YGLAVETKFLQAGAPAGEKSGALPQCVPA GENKSKXDESQ 883
D 903 NHQLRD-----HNIRP-----GNKYNVCS-----GESAIKVKKA 923
QY 884 SLLRRRRSGVFCANCLTKTSLMRKNANGVYVNCAGLYOKJLHSTPRPLNIHQKNGEQ 943
D 924 ELIK-----GNKYNVCS-----RTFSENG-- 944
QY 944 IRRRTKRLNP-----EALQABQLNKOGRNEQVNGSPLEKRSER 986
D 945 -LREHMOTHLGPVHYMCFICGERFPSSLTLEHKVTHSKSLDTGNCRICKMPLQ--SEE 1001
QY 987 HLTESHQREIPLPSLSKYEAGSLTKSHSAQPVLVSCOTLDIHKMOPHLHIQIKSPQEST 1046
D 1002 EFLEHCQMH---POL-----RNSLTGFRVCMQTV-STJEL-KIHGTFFVQ-----1044
QY 1047 GDPGNSSVSEKSGSSERGSPIEKYMRPAKPNYSPPGSPLEKYQYPLFG-PFVHNDPQS 1106
D 1045 -KTGNGSSVQ---TTGRGQHVQK-----LYKASCCLKEFRS 1076
QY 1107 EADMLRFWSKYKLSVPGNPHYLSHVPLNPNQNY--VPYTFNLPHFSVAVG-----1157
D 1077 QODLV-----KLING-LPY-----GLCAGVNLKSKSSSPGLSLPPGASRFLGQWESL 1124
QY 1158 -----SDNDIPLDLAKHSRPGPTANGASKETKAPPNVKN 1193
D 1195 SAMEGKGAGGLKTRCSSCNVKESESELQHIQTVHRELVPDANSTQLKTPQVSP---1180
QY 1194 EGPLNVNATEKVDRTQDELST-KCVHCGIVFLDE-----VMYALHM-----1234
D 1181 -----MPRISPSQSEKITYQIKCMQVYVNEWICQVVAHMKIDBGLNHECKLCSQ 1232
QY 1235 -----SCH-----GDSGPPQCSIC-----QHL-----CT 1253
D 1233 TFDSPAKLQCHLIEHSPEGMGTKFCPCVCFVFFVOANKLQCHIISAHQEDKIYDCTQCP 1292
QY 1254 DKYDFTTHIQ 1263
```

Db 1293 QKFFQTELO 1302

RESULT 7  
Q8C1Q1

Q8C1Q1 PRELIMINARY; PRT: 1271 AA.  
AC 98C1Q1  
CT 01-VAR-2003 (TRENBLrel. 23, Created)  
ET 01-VAR-2003 (TRENBLrel. 23, Last sequence update)  
DT 01-VAR-2003 (TRENBLrel. 23, Last annotation update)  
DE Early B-cell factor-associated zinc finger protein.  
GM EFPAZ.  
CS Xus musculus (Mouse).  
CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
CC Yamalia; Euthera; Rodentia; Sciurognath; Muridae; Murinae; Xus.  
CX NCST\_TaxID:10090;  
RN 1.  
PP SEQUENCE FROM N.A.  
RC STRAIN=BA13/C;  
RA Warming S., Liu P., Suzuki T., Akagi K., Lindtner S., Pavlakis G.N.,  
RA Jenkins N.A., Copeland N.G.  
RT "Ezra", a common retroviral integration site in murine B-cell lymphoma,  
RT encodes an EFPAZ-related knuppel-like zinc finger protein."  
RL Blood 2:0-3(2002).  
DR EMEL; AY:47407; XAN3940.1;  
SQ SEQUENCE 1271 AA; 14318 MW; 51E1B67F48780DFC CRC64;

Query Match 3.9%; Score 285; DS 12; Length 1271;  
Best Local Similarity 18.7%; Pred. No. 6.7e-09;  
Matches 282; Conservative 160; Mismatches 540; Indels 530; Gaps 69;

QY 26 TESKVSCKKETSADQMSNTD-----QSDAAELNKHKEHSLH-----VQD 66  
D 25 TSPALSDRNSVTSCBERKEDDEVEDESIYCDHQQDFESLADLTDRHRCPCGGDD 84  
QY 67 P-----SSSSKKDLKS--AVLSE--KAGFNYESPKGNFPSPFFHDEVTDNMIAFPFP 116  
D 85 PQLSVASSPSSKQVASTQM:GDGCDLGLGEEGGTGLPYPC---QFCDSKFLRSY- 139  
QY 117 AGGVCEPLKSPRAEADPDQMACTPGSCSLSTEDQMSPKATEETGQASQANQCG 176  
D 140 -----LKRHEQIHS-DKLFPKCTPCSRLPKHXSRDRHIK---HTGCKRYHCHCECA 188  
QY 177 LSPVSVASKNPQVPSDGGVRLNKSKTDLVNDNPPAPLSPLEQDFKQNICGYGVYKNDP 236  
D 189 AFS-----RSDHLKHLKTHSSXF-----FKCSVCKRGF--SST 221  
QY 237 TDLIKHR-----KYHJGLHNT-----RODAELSKILAJNMVQFSSHS 276  
D 222 SSLQSHVQAFKKKHENLAKSEAKKDDFCMCDYCEDTFSTOELEKRVJLJH--POLSEK 279  
QY 277 KDFQKN-----RSVFSGLQDINSR-----298  
D 280 ADLQCHCEPEVDCSTLLAH-HQAHANQKRCPCPCQPSVSEVYCHLDSHROFDDSN 339  
QY 299 -----PVLLNGTYPVQVTSQGTFFIGQEKTPDC-----QGNTK-----331  
D 340 HVSVPDPVJ-----GSVASMSSATPDSASVERGSTPSTLKLPLRGCKMKRDD 367  
QY 332 -----YPRCKFC--NFTYMGNSSTLEQHFJCTHPNK:KASLPSSEV---AKPSSEK 378  
D 388 GSWPKAVV7SCPYCKSRDFT----SLAVLEIHLKTHADKPCQS-HTCQICLDSMPTLYN 442  
QY 379 SNKSPALQCSSPS-----GDLGKWKQKIVK-AGDETTPGVYSVP 416  
D 443 LHEHVKLHKSHAYPVMQFNGNISAFHCNYCEMPADINSLQEH:RVSHCGGNA-----495  
QY 417 IKFLDSSRQNGTEATSYVMCKFCSCSSSSSLK:LEHYGKHGAVQSGGLN-----468  
D 496 -----NPDGNAFCNQCSCMGFLTES--LLEHIQAHCSVST:KLESPPVQCTO 544  
QY 469 -----PELNKLSRGSVI-----NQNDJ-----AKSSSEGETYTKDKSSSGAKK 507

```
Db 545 SPFEVSCPTNSPIFGSLKLTUKHKNKIP:AHKKSKAEOSPVSSDEVSSPR 604
QY 508 KDFSSGAEDNMVTS--YNCQFCDFPRYSGHGDVIVWGLRHYYQQLHN:HKCT:KHCP 565
Db 605 QRLSGSA--NSISNGEYPCNQCDLAFSNFES-----FQTHLKLHLELLRKQACP 652
QY 566 FCPRGILCSPEKHGEITYPACRKN--CSHC-----ALLLHLSPGAGSSR 611
Db 653 QCKEDFQSGESLLOHLTHVMTTSTHYVCESCDQFSSVDDLOKHLLDMH-----TF 704
QY 612 VRHOCHOCSTTPDQVLLPHYESVHESQAS-----DYKQEAHLQSGDQGSVE 662
Db 705 VLYHCTLCEVDFSKUSIQVHLAVKSNKMYRCTACWDFRKA-----DLQVHVKH 758
QY 663 S-----KHSCTKCFITQVEEISRHVRAHSCYKQCQSFADTOSLSEHFNTHVC 716
Db 759 SHLGNPAKAHKCFCGETESTEVELOCHITTHSKKYNCRFCSPKAFHAVILLEKHLREKHC 818
QY 717 QEQDITANGEEDEGHAISTIKKEPKIDFRVYNLLTSPSKMGEPVSVVKREKLEKGL 776
Db 819 -VFDAANGTANGVPPTSTTKAEPADLOGLMLKNPEA-----PNSH-----859
QY 777 KEKVMTSESSDDL-RNVTVRGADILRGSPSYTOASIGLLTPVSGTQEQTKLRDSPNEA 835
Db 860 -----EASEDDVASEPMYGCDC--GAAYTMEV-----LQHR-LD-----895
QY 836 AHLARIYGLAVETKQFLOGAPAGGKSGALPOQYPASGENSKDSQSLRRRGSGVF 895
Db 896 -HNIRP-----GEDGSRKAEFIKSHKNCVCSRTF-----F 927
QY 896 CANCLITKTSLEKKNANGGVVNCAGLYOKLHSTPRPLNI:KQNGEQIIRTRKRLNP 955
Db 928 SENGLEHLOTHRGPAK-HVMCP:CG-----ERFPLLT:TBH-----964
QY 956 EALQAEQLKQKRGSEEEVQNGSPLRRSEDLHTEHQREIPLPSLSKYEAQGSJ--KS 1013
Db 965 -----KVTHSKSLDTGCRICKXPLO--SEEEFIEHCQM--PDJ--RNSL:GPRC 1009
QY 1014 HSAQQPVLVSQTLDIHK--RMQPLHIOIKSPQESTGDPNGSSSVSEKSGSSEKSPYK 1072
Db 1010 VVCMQTVTSTLELKHGTTHNQKL-----AGSSAAS-----1040
QY 1072 MRPAKHPNYPSPGSPTEKYQPLFGLPFVHNDQFSEADMLRFRNSKYLSVGNPHYL--- 1026
Db 1042 -----SPNGQGLQK-----LYKALCLKEFRSKQDLVR-----JDNGLPYGJGAG 1061
QY 1129 -----SHVPGGL--PNPCQNYVPYPTFNLPHPFSAVGSDNDIPLDLAKHSRPGPTANG 1079
Db 1082 CMARSANGQVGLAPPEPADR--PCAGLRCPCECNVXFESAED:ESHKQYDHRCLTPE:SG 1039
QY 1180 ASKEKTKAP-PNVKNGPLNVVTKVDKSTQ-----BELSPKCVHCG:VFLDEVM 1229
Db 1140 PRGAQTSVPVRKTYQCIKCRMTFENEREIOIHVANHMIEGINHECKLNCMPDSPAK 1199
QY 1230 YALHM--SCHGDSGPRQCSIC-----QHL-----CFDXVDFTT 1260
Db 1200 LLCHLIEHFEGMGITKPCVCFVTVQANKUQCHIFAVHGQEDKXYDCSQCPQFFPOT 1259
QY 1261 HIQRLGLHRNNAQ 1272
Db 1260 ELQNHTMSQHAQ 1271
```

## RESULT 8

```
Q8VDS6
ID Q8VDS6 PRELIMINARY; PRT; 1311 AA.
AC Q8VDS6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2003 (TrEMBLrel. 20, Last sequence update)
DE Hypothetical 147.7 kDa protein.
OS Mus musculus (Mouse).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 1;
RP SEQUENCE FROM N.A.
RA Strausberg R.
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EXBL; BC021376; A4421376.1; -.
DR InterPro; IPR00345; CyC_heme_bind.
DR InterPro; IPR07087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 28.
DR SMART; SM00355; Znf_C2H2; 30.
DR PROSITE; PS0190; CYTOCHROME_C; 2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 27.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 22.
KW Hypothetical protein; Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 1311 AA; 147665 MW; C8D92B5335C099A CRC64;

Query Match 3.9%; Score 264.5; DB 11; Length 1311;
Best Local Similarity 19.2%; Pred. No. 7.6e-09;
Matches 302; Conservative 177; Mismatches 516; Indels 575; Gaps 82;

QY 1 MVRKKNPPLRVVASEGEGQILEPIGTESKYSKKNKFSADQMSQNTDQS-DAA-----EL 54
Db 1 MSRRKQKPRSL-KDPNCKLEDKIEDSEAVCKKRPEDGGELEDAVHSCDSCLOVFESL 59
QY 55 NKKEEHSH-----VQD-----PSSSKKDLKSAVLSEKAGFNYESPSKGNFP 98
Db 60 SDITEHIIHQQLTGDVDEVEDPCSPASSPSKQDQTSFSGEGCDPGEB---EGG-P 114
QY 99 SFPHD-SVTDNMLAFSPFPAAGGVEPLKSPQRAEADPDQDMACTPSGDSLETKEQKYS 157
Db 115 GLPYPCFCDFKSFRLSY-----LKHQEGSHS-DKLPFKCTYCSRLFKHKSRRDH 164
QY 158 PKATEETGCAQAGGAGNCG-----LSPVSVASKNP---QVPSGGVRLNKEKTDLL 205
Db 165 IKL-HGDKKYH-SECDAAFSRSCHLKIKHTSKNPKYKCAVCRGFLSSSS-HGHQ 222
QY 206 VND-NPDPAP-SPPELQDF-----KCNIGYGYGNDPTDLIKHFRKYH-----247
Db 223 VHEANKDGSOGSRMEDKMKMDCKCKSQCEGF--DFPELQKH:AECHPSCSPNEDEAA 280
QY 249 ---LGLNRTQDAELFESKILALNMVQ-----FSHKDPQKVNRSVF 287
Db 281 LQMYCHELFVEETS-LNHHIEQVHGSGKKNSCS:CSSEF:TVVELYSHMDSHQPESC-- 338
QY 298 SGV-LQDINSRPVJLN-GTYDVQVTSQGTFIGI-----GRK-----TPD 325
Db 339 -----NHSKSPSVTVGYTSVSSTTFCSNLVSDSSTMVZAAPIPKSRGRKRAAQOTSC 392
QY 326 CQGNK-----YFCKKFCNFTYKNSSTELQHPLOTH---PNKIKASLSPSEVAKPSEK 377
Db 393 MTGPSSKQAKVTYSCLYCN-KQ:FPSSLAIV:QHLKTMHLCKPEQAH:COYCLEV-PSLY 450
QY 376 MSNKS:PAL-QSSSG-----DLQKQCKITVRAGDCTP-----410
Db 451 KLNELKQVHEAQPGG--VSAPVAVVQCNFCSEVNDLNTLQEHTRCSRSEGFANPAKDS 510
QY 411 -----VQYSVPIKPLDSSRN-----GT---EATSYWCKFC 439
Db 511 NAFFCPHCYNGFLT-----DSSLEEHIRQVHCD:SGSRFPSPVLGTKEPWEVYSCSYC 565
QY 440 SPSCESSSLKLEHYKQKHGAVQSGGLNPELNDKLSRG-----SVINQDL-----486
Db 566 TNSPIFNVLKLNKH:KENHKNIP:ALNYIHNGKKSRLSPSPVA:EOTT:KQMQTVG 624
QY 497 ---AKSSEGETMTKTDKSSSGCAKKDFSS-----KGAEDNMTSYNCCQCFDPRYSKSHGP 538
Db 625 GGPARRAS-GEYIC-----KQCGAKYFSLDSFQTHLKTHLDTVLPKLTCPQCKEPNQES- 678
QY 639 DVIWVGPLRHYYQQLHN:HKCT:KHCP:CFPRGLCSP---EKHLGEI-TYF:ACKSKNSCH 594
Db 679 -----LLXKV-TIMEKINTSYIYICESCDKQFTSVDCLOKHLDMHTFVF-----FRCTL 726
```



```

Db 773 SSFDS-IETPALVAERE-----:|||:|PVEKLASEPPRN 601
Qy 934 NIHKQNGEQIIRRRTRKRLN-DEALQAEOLNKQQRGSNEEQVNGSPLERRSE-----985
Db 802 PLISEEAPNTFFKAALTAETVPLPPFPESESLLKAYRRQDQKECAELVLSGRVQMVIQGE 962
Qy 986 -----DHLTESHQREIPLPSLSKYRAQGSLLTKSHSAQOPVLVSQTLDIHKMQPL--HI 1037
Db 862 GRAFRCPHCFPIIRREKALTTHSKSCQG-----RREPLLCPEGCAGFKOORGLSTHM 914
Qy 1038 QIKSPQESTGDCPNSSSVSGKSSRGSPLEKYMRPAKHP-----NYSPPGS 1085
Db 915 MKKCPV-----LLKNKALPKPVSPTLHPQLPDNQASQDAESRKPPPL 957
Qy 1086 PIEKYQVPLFGJPFVHNDFOSEADWLRFWMSKYKLSVPGN-PHYLSHVPGJPNPCQNYVPY 1144
Db 958 P-----SKVELLPKDAFSDLPGGPGVVEEPLPTFSDF 989
Qy 1145 PTNLPHPFSVGSNDNDIPLDLAIKHSRPOGTANGASKETKAPPNVQNEGLPNVVKTEK 1204
Db 990 PT-SPP-----ENSLP-----TGTSK-----1005
Qy 1205 VDRSTQDELSTKCVHCGIVFLDEVMYALHMSCHGDSGPFQCSICQHLCTDKYQFTTHIQR 1264
Db 1006 -----FHFQCKFHCSCTFLCSRLSLISISHTVE 1034
Qy 1265 GLHRNNAQVKNKPK 1280
Db 1035 GCRGGQKRGKRGKRPQ 1050

RESULT 10
Q8UVK1 ID Q8UVK1 PRELIMINARY; PRT: 589 AA.
AC Q8UVK1;
DT 01-MAR-2002 (TEMBLrel. 20, Created)
DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Ikaros-like transcription factor IKLF1.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hypercartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OC NCBI_TaxID=7757;
RN [1]
RE SEQUENCE FROM N.A.
RA Mayer W.E., O'Ruigin C., Terric J., Saraga-Babic M., Tichy H.;
RT "Identification of two Ikaros-like transcription factors in lamprey.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF424733; AAL67302.1;
DR InterPro: IPR007087; Znf_C2H2.
DR Pfam: PF00096; Znf_C2H2_5.
DR SMART: SMC0355; Znf_C2H2; 6.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE: PSS0157; ZINC_FINGER_C2H2_2; 4.
KW Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 589 AA; 63886 MW; 90C00A5C3167F61A CRC64;

Query Match
Best Local Similarity 19.4%; Pred.NC.3.2e-09;
Matches 143; Conservative 62; Mismatches 190; Indels 343; Gaps 29;

Qy 603 SPG-----AAGSSRVKHQCQCSFTTDPDVLPHVSVHESQASDVKO-EANKHQSGD 656
Db 117 SPGETRFSSGSSKLAADVCGMVICGNVLVWVKRSHTGTPFQCSQCGASFTQXGL 176
Qy 657 QQSVK-----ESKHSCTKCFITQVEEISRYRAHSC---YKRCQCSFTAAQTQSLBH 710
Db 177 VRHKLHTDEKPFKCHLCSACRRDALMGLH-RTHSVGPKYKCHSCRCYKORSLEH 235
Qy 711 FNTVHQEQDITTTANGEDGHAISTIKEPKIDFRVNLNLTPOSKVGPVSVKREK 770
Db 236 LERCPAYCQQLSTRNQEDA-----DIRVH-----MDQPV-----266

```

```

Qy 77: BEKDCLKEKVMWTESSDOLRNVTWRGADILRGSPSYTOASLGILLTPVSGTQECQTKTLRDS 830
Db 267 ---DGLLE-----271
Qy 831 PNVEAAHLARPIYGLAVETKGFQGAAGAGKESGALFQCYPPASGENKSKDESOLLARRR 890
Db 272 ---AGSDRI-PLPNQLPG-----SFLAKRK 292
Qy 851 GSGVFCANCLTTKTSARKNANGVCNAGLYQKJHSTPRPLNKKONNGEQIIRRRTR 950
Db 293 S-----STPQKL-----FNQ 302
Qy 951 KRLLPEALQAEALQKQKRGNEBQVNSPLERSEDLTSHQREIPLPSLSKYKQGSLL 1010
Db 303 KRLN---NCLSDIRCEQWSSDO--NEAP-----HFVEQ-----AMSTYLYGYRPL 342
Qy 1011 TKSHSAQOPVLVSQTLDIHKMQPLH-QIKSPQESTGDCPNSSSVSGKSSRGSPLEK 1070
Db 343 LSQSPSQPLVLSHSES--NRIOF-----PSLSLAIP-----372
Qy 1071 YMRDAPKHPNPPSPGSPLEKYQVPLFGJPFVHNDFOSEADWLRFWMSKYKLSVPGNPHYLSH 1130
Db 373 -----KFPNS-IEDYA-PVIGAVY-----SH 392
Qy 1131 VGPLNPNQNVVPTTNLPPH-PSAVG-----SDNDIPLDLAIKHSRPOGTANGASKEX 1184
Db 392 TVGQSPSPRSQSP-----LPCHGFCSSQKQGTSEQLPRDRASLGFHGRAAAASVSSPS 446
Qy 1185 TKAPPNVKE-----GPNVYVKEK---VDNSTCQELST-----1215
Db 447 NSCPSTSTESSHEERORLFRANCSAQPTGNSSTQCAASTPARQDVQLQEGGG 506
Qy 1216 -----KCVKCGIVFLDEVMYALHMSCHGDSGPFQCSICQ 1249
Db 507 QPRPPASGGLAVRMGEGDGEYMRAYNCLHCQVIFLDHVMVTLHMGCHGRDQPFECNCG 566
Qy 1250 HCTCKYDFTHICRGJH 1267
Db 567 HRSRDYEFSSHIRGEH 584

RESULT 1:
Q8ESD2 ID Q8ESD2 PRELIMINARY; PRT: 1167 AA.
AC Q8ESD2;
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Early B-cell factor associated zinc finger transcription factor.
GR EBF2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RE SEQUENCE FROM N.A.
RA Croci L., Corradi A., Vauti F., Wurst W., Koochi N., Gonzalez G.G.;
RT "cDNA sequence and map assignment of Ebf2, orthologous to the zinc
RL finger transcription factor gene Roaz."
DR EMBL: AF188609; AAG17053.1;
DR HSSP: P08347; 1SP1
DR MGC: MGC1891217; Ebf2.
DR InterPro: IPR007087; Znf_C2H2.
DR Pfam: PF00096; Znf_C2H2; 28.
DR ProDom: PD000003; Znf_C2H2; 2.
DR SMART: SM00355; Znf_C2H2; 26.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 27.
DR PROSITE: PSS0157; ZINC_FINGER_C2H2_2; 21.
KW Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 1167 AA; 131578 MW; 6EB8FA77-E0CDF3E CRC64;

```

Query Match 3.8%; Score 262; DB 11; Length 1167;

Best Local Similarity 19.1%; Pred. No. 9.5e-09;

Matches 261; Conservative 138; Mismatches 512; Indels 456; Gaps 60;

```
QY 27 ESKVSKNKEPSADQVSENTDQSDAAELNHKEHS:HVQCPSSSSKKDL-----75
DB 136 KSEAKKDDFCDCYCEDTFSQTEELE---KHVLTLLH---POLSSKADLQCHQPEVEFD 189
QY 76 KSAVLSE---KAGFNYESPSKGNFPSP-----PHDEVTDKMLAFSPFAAGGYCE 123
DB 190 ESTLLAHIHQANQKXKCPMCPQEPQESSVEGYCHLDHRQPDSSNHSVSPDPVGSVAS 249
QY 124 PLKSPORAEADPDQDMACTPSSGDSLETKEQKMSPRATEETQCAQSGOANQCG-SPVSWA 183
DB 250 ---MSSATPDSSASVERGSTPDTLKLPGCKKMRCDQSM-----287
QY 184 SKXPQVPSDGGVRLNKSRTDLLV-----NENPDPAPLSPQELDFKCNICGYYGN 234
DB 288 ---PKVVSYPYCSKRDFTSLAVLE:HLKTIHADRPQ-----QSHTCQIC---LDS 332
QY 235 DPT--DLIKHFRKYHLGLHNRTRQDAELDSKILALNNVQFQSHSKDFQKVRNVSFSGVLQ 292
DB 333 MPTLYNLNHRVKLHKS-----HAYPVMQFCNISAP---HCNYCPSEYFA 373
QY 293 DINSRPVLLNGTYOVVTSGGTFTIGIKRTPDQGN:TKYPRCKFCNFTYMGNSTELEQ 352
DB 374 DINS-----LQEHIRVSHCPNANPPD--GNNAFF--CNQCSMGFLTESS--LTE 417
QY 353 HF-QTHPNKIKASLPSESVAKSEKSNKSI:PALQSSDGLGKQKQDKITVKAQGDTPVG 412
DB 418 H:QOAH-----CSVGSKLESP-----434
QY 413 YSVPIKPLDSSQNGTEATSYNWKFCSPSCSSSLKLLHYGKHQGAVGSGG:NPESLN 472
DB 435 ---VQOPTQSFME-----VYSCPCTNSPIFGSILK:TKHIKENHKKI-----PLAH 478
QY 473 DKLSRGSVINQDLAKSSGE:WTKTDKSSGAKKQKPFSSKGAENMVTs---YNCQCFDF 530
DB 479 SK-----KSKAEQSPVSDVEVSFKQRLSGSA---NSISNGSYPCNQCTL 522
QY 531 RYKSKSGPDVIVVGLRHYQQLHNHKTIXHCPFCPRGLCSPEKH:GEITYPACRKS 590
DB 523 KFSNFES-----FQTHKLHLELLRLKQACPOCKEDFSQESL:QHLTVHWT-ST 573
QY 591 N--CSHC-----ALLLHLSPGAAGSSRVHCHQCSFTTPDVVLLHFHVESV 636
DB 574 HYVCESCDXQFSDVDLQKHLDMR-----TFVLYHCTLCQKVFOSKVS:QVFEJAVK 625
QY 637 HESCAS-----DVKQEAHLQSGDQGSVKES-----KHSCT:KCDFTTCQVEEI 681
DB 626 HSNKKNYKRTACNWDPRKEA-----DLQHVKHSHLGNPAKAHKC:FCGFTFTSEVEL 679
QY 682 SRHYRHAHSYKCRQCSFTTAADTQSILLEHFNTHVCOEDIT:ANGEEDEGHASTIKEPK 741
DB 680 QCHITTHSKYKNCRCFSKAPHAVILLEKHLREKHC-VFADAENGANGVPPSTKKAAP 738
QY 742 IDPRVNVLLTPDSKMGEPVSESVVREKLEKDKLEKVTWTESSDDL-RNVTWRGAIL 803
DB 739 ADLQGMLLKNPEA---PNSH-----EASEDDVJASEPMYGCIC 774
QY 801 RGSFSYVQASLGELTPVSGTQEQTKTLRSPNVEAAH:AEPPVGLAVERKGLQAPAGG 860
DB 775 --GAATYMEVL-----LNHRLRD-----HNIRP-----G 797
QY 861 EKSGLAPQOYPASGENSKDESQSLRRRRRGSGVFCANCLITKSLWRKNANGGYVKNAC 920
DB 798 EDDGSRKKAZF:KGSKKNVCSRTF-----PSENGLREHLQTHRGPAK-HVMCP:IC 847
QY 921 GLYQKHLSTPRPLNI:KONNGEOI:IRRTKRLNPEALQAEQ:LNKOORSGNEEQNGSPL 980
DB 848 G-----ERFPSLLTLEH-----KVTHSKSLDTGTCTICKMPL 860
QY 981 ERSSEDLHTESHQREIPLPSLSKYEAGSILT--KSHSAQPPVLVSQTLDIHK--RYQPLH: 1036
```

## RESULT 12

```
Q925J4 PRELIMINARY; PRT: 1669 AA.
AC Q925J4;
DT 01-DEC-2001 (ITEMB:rel. 19, Created;
DT 01-DEC-2001 (ITEMB:rel. 19, Last sequence update;
DT 01-MAR-2003 (ITEMB:rel. 23, Last annotation update)
DE Zinc finger 142.
GN ZFP142 OR ZNF142.
OS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
CX NCB: TaxID=10096;
RN 11;
RP SEQUENCE FROM N.A.
RC STRAIN=ILS;
RA Bringer M.A., Thompson J., Conroy C., Xu Y., Yang F., Canniff J.,
RA Beeson M., Gordon L., Bennett B., Johnson T.E., Sikela C.M.;
RT "High-Throughput Sequence Identification of Gene Coding Variants
RT within Alcohol-Related QTLs";
RJ Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EXBL; AF132092; AA056118.1;
DR MSD; MG11924514; Zfp:42.
DR InterPro; IPR003045; Cytochrome_bund.
DR InterPro; IPR003045; Cytochrome_bund.
DR InterPro; IPR003045; Cytochrome_bund.
DR Pfam; PF00096; Zf-C2H2; 30.
DR SMART; SM00355; Zf-C2H2; 36.
DR PROSITE; PS0180; CYTOCHROME_C_1.
DR PROSITE; PS02290; ZF_MHC_1.
DR PROSITE; PS00026; ZINC_FINGER_C2H2_1; 17.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 19.
KW Metal-binding; Zinc; Zinc-finger.
SC SEQUENCE 1669 AA; 186763 MW; 7D5138424A5198F6 CRC64;
```

## Query Match

Best Local Similarity 18.4%; Pred. No. 1.7e-08;  
Matches 213; Conservative 124; Mismatches 360; Indels 459; Gaps 50;

```
QY 366 LPSEV-----AKPSEKSNKSI:PALQSSDGLGKQCK:TVKAGDTPVGVSV 415
DB 113 LPSQELLPAKLPFGHREPEEAST-PLPQESABEED-ABEESVTQDSQKV-MDKSQ 169
QY 416 PIKPLDSSRQNGTEAT-SYWKCKFC-----SFSCEs-----SSS 448
DB 170 GACQCEHGVSGTGESLFKTHYCPCKRCFKKXTHLVSH:HLHPPDPSLQCPCKKFFTSK 229
QY 449 LKLEHYGKHGAVCGSGLNPELNDKLSGVS:NQNDJAKSSEGETMTYTKSSSGAKKK 508
```







QY 231 YGNDPTOLIKFKRYHGLGHNTRODA-----ELDSKILALHNMVCFSHSKDFOKVN 283  
D5 255 F--DFPEDJOKHIAECHPECSNEDRAALQCYCHLFEETSLNMHKEOVHSGS-KKNS 311  
QY 284 RSVFSGVLODI-----NSSRPVLLN-QTYDVQVTSGGTF:GI----- 319  
D5 312 CS:CSSEFHTVEELYSHMDSHQPESCNHSNPSLVTVGYTSVSSTPDSNLSVDSSTMV 371  
QY 320 -----GRK-----TPDCGNTK-----YFRCKPCNFYMGNSSTELQOHLQTH- 358  
D5 372 EAAPPKPSGRKRAAQQTDMTPGSSKQAKYTSIYCN-KOLFSSLAVLQIHLKTMHL 430  
QY 359 --PNKIKASLPSSEVAKPSKGNKSTPAL-QSSDSG-----DL 394  
D5 431 DKPEQAHICQYCLEVL-PSLYNNEHLKQVHEAQDPGLIVSAMPALVYQCNFCSEVWDL 489  
QY 395 GKQDQKITVAGDDTP-----VGYSVFKPLDSSRQN----- 426  
D5 490 NTLQEHIRCSHGFPANPAAKDSNAFFCHCYMGFLT-----DSLEBHIRQVHCHDLSGRF 544  
QY 427 -----GT---EATSYWCKFCFSCSSSSLKLEHYGKHQGAQVQSGGLNPELNDKLSG 478  
D5 545 GSPVLGTPKEPVVEVYSCYCTNSPIFNSVLKUNKHIKENHKNIP-ALNYIHNGKXRA 603  
QY 479 SVINQNDLAKSSEGETMTKTDKSSGAKKXDFSSKGAEDNMVTSYNCQCFDFYS----- 533  
D5 604 ---LSPLSPVAIBQTSLKQWQAVGGAPAP-----TGEYICNQCQAKYTSJRSF 649  
QY 534 ---KSHGPDVI---VVGPD-----LLRHYQQLHNIHCTIKHCFPCPRGLCSP-- 574  
D5 650 QTHLKHLDVLPKLTCPQCNKFPNQESLKHV-TIHFMTITSYVYICESCDKQFTSVDO 708  
QY 575 -EXHLGEI-TYPPACRKSNSHC-----ALLLHLSGAAGSSRVKQCHQCSFTT2D 625  
D5 709 LQKHLLDMHTFVP-----FRCTLQEVFDSKVISQLHLAVKHSNEKV-YRCTSKWDFRN 763  
QY 626 VDULLFHYESVHESQASDVQOEAHLQSGDQOSVKESHEHCTKC--DFITQVEEE-- 682  
D5 764 ETDLOLH-----VXH--NHLE-----NQGKVKHC:FCGESFGTEVELQCHI 802  
QY 682 SRVYRAHSCVKFCQCSFTAAJOTOSLJEHFTVHCQEODITTANGEDGHAISTIKKEPK 741  
D5 803 TTHSKK-----YNCKFCXAFHAIILEKHUREKHC-VFETKTFNGTNGASEQVQKEVE 857  
QY 742 IDFRVYNLLTPDSKXGEPVSESVVYREKLEEKQGLKEKWWTSSDDLXNVWAGADILE 801  
D5 958 LQ-----TLITN-----SQSSHNSHDGSEEDVDTSE-----PMYGCDC- 891  
QY 802 GSPSYTOASLGLLTPVSGTQEQ:KTLSDSPNVEAHLAREYGLAVETKFLQGFAGSE 862  
D5 892 -GAAYTMETL-----LQNHQLED-----HNRP----- 913  
QY 862 KSGALPOQYPASGNKSKDESQSLRRRRGSGVFCANCLTTKTLMRKNANGYVCNAG 921  
D5 914 -----GESAVKKKAEI:K-----GNYKCNVCS 936  
QY 922 LYQKLHSTPRPLNIKONNGEQIIRTRKRLNP-----EALQAEQLN 964  
D5 937 -----RTFFSENG-----LREHMQHLGPHVHYMCPICGERFPPSLTLTEHKVTH 982  
QY 965 KQORGSNEQVNGSPLERRSEDLHTESTHOREIP:PSLSKYEAOGSLTKSHAQOPVLVSC 1024  
D5 982 SKSLDTGNCR:CKMPLQ--SEEEFLEHCOMH---PDL-----RNSLTGPRCVVCMQTVTS 1031  
QY 1025 TLDIHKRMQPLHIQ:KSPQSESTGDPGNSSSVSEKGSSEKSGSPIEKYMRPAKH2NYSPPG 1084  
D5 1032 TLEL-KIHGTFHMQ-----KTNGSAVQ-----TTGRGQHVKX----- 1063  
QY 1085 SPIEKYQVPLPGLPFVNDPOSEADHLRFPWSKYKLSVPGNPHYLSHVPLPNCQY--V 1142  
D5 1064 -----LYKASCULKEFRSKQDLV-----KLDINGLPY-----GLCAGCVNLSKS 1102

QY 1143 PYPTENLPP-----HFSAV-----GSONCIPDLAIKHS 1171  
D5 1103 ASFGINVPOTNEPG:GQENLSAIEGKGVGLXTRCSCNKKVFESESELQNHICTIHR 1162  
QY 1172 RPSGTANGASKKXKTRAPPNVKNBGP:NVVTKVKVRSTODEJST-KCVHCGIVFLDE--- 1227  
D5 1163 ELVPSDSNSTQLATFOVSP-----MPR:SPSQSDEKKTQCIKQCMVFNEMDI 1210  
QY 1228 -VYALHLM-----SCH-----GDSGPFQCS:G----- 1248  
D5 1211 QVHVANMIDEGLNHECKLCSQTFDSPAQLQCHLIEHSEFGMGSTFKCPVCFVVFQANK 1270  
QY 1249 --QHL-----CTDKYDFTTH:Q 1263  
D5 1271 LQOHIFSAHQEDK:YDCTOCQPKPFFOTELQ 1302

Search completed: October 29, 2003, 11:42:49  
Job time : 199 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 29, 2003, 11:44:12 ; Search time 236 Seconds  
(without alignments)  
8418.074 Million cell updates/sec

Title: US-09-702-216-1\_COPY\_500\_5000  
Perfect score: 4501  
Sequence: 1 agctctctgtgtctctga.....agaaaaatggcgccctctca 4501

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Capext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:  
1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:  
2: /cgn2\_6/prodata/2/ina/5B\_COMB.seq:  
3: /cgn2\_6/prodata/2/ina/6A\_COMB.seq:  
4: /cgn2\_6/prodata/2/ina/6B\_COMB.seq:  
5: /cgn2\_6/prodata/2/ina/6C\_COMB.seq:  
6: /cgn2\_6/prodata/2/ina/backfileseq.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	65	1.4	1456	3	US-09-037-135-1
2	58.6	1.3	1984	4	US-08-733-622C-1
3	57.4	1.3	1004	1	US-08-465-590-8
4	57.4	1.3	1004	3	US-08-283-300A-7
5	57.4	1.3	1004	3	US-08-711-417C-8
6	57.4	1.3	1004	4	US-08-733-622C-21
7	57.4	1.3	1004	5	PCT-US95-09345-7
8	57.4	1.3	1386	3	US-08-465-590-3
9	57.4	1.3	1386	3	US-08-283-300A-2
10	57.4	1.3	1386	3	US-08-711-417C-3
11	57.4	1.3	1386	4	US-08-733-622C-16
12	57.4	1.3	1386	5	PCT-US95-09345-2
13	57.4	1.3	1551	3	US-08-711-417C-165
14	57.4	1.3	1611	5	PCT-US93-08743-3
15	57.4	1.3	1788	4	US-09-435-327A-19
16	57.4	1.3	3629	4	US-09-435-327A-18
17	53.8	1.2	2897	2	US-08-927-394-1
18	53.8	1.2	2897	4	US-09-016-434-1163
19	53.2	1.2	7218	1	US-08-232-463-14
20	50.6	1.1	1128	1	US-08-465-590-7
21	50.6	1.1	1128	3	US-08-283-300A-6
22	50.6	1.1	1128	3	US-08-711-417C-7
23	50.6	1.1	1128	4	US-08-733-622C-20
24	50.6	1.1	1128	5	PCT-US95-09345-6
25	50.6	1.1	1170	1	US-08-465-590-6
26	50.6	1.1	1170	3	US-08-283-300A-5
27	50.6	1.1	1170	3	US-08-711-417C-6

28	50.6	1.1	1170	4	US-08-733-622C-19
29	50.6	1.1	1170	5	PCT-US95-09345-5
30	50.6	1.1	1296	1	US-08-465-590-4
31	50.6	1.1	1296	3	US-03-283-300A-3
32	50.6	1.1	1296	3	US-08-711-417C-4
33	50.6	1.1	1296	4	US-08-733-622C-17
34	50.6	1.1	1296	5	PCT-US95-09345-3
35	50.6	1.1	1788	1	US-08-465-590-2
36	50.6	1.1	1788	3	US-08-283-300A-1
37	50.6	1.1	1788	3	US-08-711-417C-2
38	50.6	1.1	1788	4	US-08-733-622C-15
39	50.6	1.1	1788	5	PCT-US93-08743-2
40	50.6	1.1	1788	5	PCT-US95-09345-1
41	50.6	1.1	2049	1	US-08-465-590-5
42	50.6	1.1	2049	3	US-08-283-300A-4
43	50.6	1.1	2049	3	US-08-711-417C-5
44	50.6	1.1	2049	4	US-08-733-622C-18
45	50.6	1.1	2049	5	PCT-US95-09345-4

ALIGNMENTS

RESULT 1  
US-09-037-135-1  
; Sequence 1, Application US/09037135  
; Patent No. 6,07034  
; GENERAL INFORMATION:  
; APPLICANT: Weigel, Ronald  
; TITLE OF INVENTION: GATA-3 EXPRESSION IN HUMAN BREAST  
; TITLE OF INVENTION: CARCINOMA  
; FILE REFERENCE: SCR-76P  
; CURRENT APPLICATION NUMBER: US/09/037,135  
; EARLIER FILING DATE: 1998-03-09  
; EARLIER APPLICATION NUMBER: 09/037,135  
; EARLIER FILING DATE: 1998-03-09  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 1456  
; TYPE: DNA  
; ORGANISM: H. sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (120)...(1451):  
US-09-037-135-1  
Query Match 1.4% Score 65; DB 3; Length 1456;  
Best Local Similarity 57.7%; Pred No. 2e-09;  
Matches 16; Conservative 0; Mismatches 85; Indels C; Gaps C;  
QV 2759 GGAGAAAACAGTCCAGGATGAATCCAGTCCCTGTACGGAGGCGTAGAGCTCCGGT 2819  
DB 1005 GGACAGAACCGGCCCTCATTAAGCCCAAGCGAGGCTGTCTGCACCCAGGAGCAGGG 1564  
QV 2819 GTTTTGTGGCAATTCCTGACCAACAGACTCTCTCTGGGGAAGATCAATGCG 2878  
DB 1665 ACCTCTGTGGCACTGTGACACCCACCAACCACTCTGGAGGAGGATGCCAATGG 1124  
QV 2879 GGATATGTATGCAACGGGTGTGGCCTCTACAGAGCTTCACCTCGACTCCAGGCTTTA 2938  
DB 1125 GACCTCTGTGCAATGCTGTGGGCTCTACTACAAGCTTCACAAATTAACAGACCCCTG 1184  
QV 2939 AACATCATTTAAACAACAAC 2959  
DB 1185 ACTATGAAGAAGGAGGCATC 1205  
RESULT 2  
US-08-733-622C-1  
; Sequence 1, Application US/08733622C  
; Patent No. 6528634  
; GENERAL INFORMATION:

```

; APPLICANT: Georgopoulos, Katia
; TITLE OF INVENTION: AIOLOS GENE
; FILE REFERENCE: 10287-030001
; CURRENT FILING DATE: 1996-10-17
; PRIOR APPLICATION NUMBER: US/08/733,622C
; CURRENT FILING DATE: 1996-05-14
; PRIOR FILING DATE: 1996-05-14
; PRIOR APPLICATION NUMBER: US 60/017,646
; PRIOR FILING DATE: 1995-10-18
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1984
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (374)...(1894)
US-03-733-622C-1

Query Match      1.3%   Score 58.6;   DB 4;   Length 1984;
Best Local Similarity 60.2%;   Pred. No. 2.5e-07;
Matches 97;   Conservative 0;   Mismatches 64;   Indels 0;   Gaps 0;

QY 3787 ATGTG-GCACTGGGATGTTCTTTCTGGATGAAGTGTATGCTTTGCAATATGAGTTG 3946
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1726 ATGTGACCATGCCAGCTCTCTTCTTAGATTATGATGTTTCAACATCCACATGGGGTG 1785

QY 3847 CCATGTGTACAGTGGACCTTTCCAGTGCAGCATGCCAGCATCTTTCACGGACAAATA 3906
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1786 CCATGTTTCCGTGATCCCTTTGAGTGTAACTGTGTGGCTATCGAAGCCACGATCGGCTA 1845

QY 3907 TGACTTCACAACACATATCCAGAGGGGCGCTTCGATAGGAACA 3947
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1846 TGAGTTCTCTTCACATGCCAGAGGAGGACAGACAGACCA 1886

```

```

RESULT 3
US-08-465-590-8
; Sequence 8, Application US/08465590
; Patent No. 5824770
; GENERAL INFORMATION:
; APPLICANT: Georgopoulos, Katia A.
; TITLE OF INVENTION: INAROS: A 1 CELL PATHWAY REGULATORY GENE
; NUMBER OF SEQUENCES: 164
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, Suite 510
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,590
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/238,212
; FILING DATE: 02-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/121,438
; FILING DATE: 14-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/946,233
; FILING DATE: 14-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Paul L.
; REGISTRATION NUMBER: 35,695
; REFERENCE/DOCKET NUMBER: MSP-027
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 7:

```

```

; REFERENCE/DOCKET NUMBER: WFG-006C2DV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1...1002
US-08-465-590-8

Query Match      1.3%   Score 57.4;   DB 1;   Length 1004;
Best Local Similarity 60.6%;   Pred. No. 3.5e-07;
Matches 94;   Conservative 0;   Mismatches 61;   Indels 0;   Gaps 0;

QY 3785 AAATGTGTCACTGGGCAATTTCTTCGATGAAGTGTATGCTTTGCAATATGAGT 3844
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 832 AGTTCGACACACTGGCGGTCTCTTCTCGATCAGTCATGTACACATCCACATGGGC 891

QY 3845 TCCATGTGTGACAGTGGACCTTTCCAGTGCAGCATATCCAGCATCTTTGCAAGCAAAA 3904
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 892 TCCACAGGTTCCGTGATCCCTTTGAGTGCACATGTGCGCTACACAGCCAGSACGGG 951

QY 3905 TATGATTTCAACAACATATCCAGAGGGGCGCTCGA 3939
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 952 TACGAGTTCTCTCCACATACACGCGAGGGGAGCA 986

```

```

RESULT 4
US-08-283-300A-7
; Sequence 7, Application US/08283300A
; Patent No. 6122279
; GENERAL INFORMATION:
; APPLICANT: Georgopoulos, Katia A.
; TITLE OF INVENTION: KAROS TRANSGENIC CELLS AND ANIMALS
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, Suite 510
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/283,300A
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/238,212
; FILING DATE: 02-MAY-94
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/121,438
; FILING DATE: 14-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/946,233
; FILING DATE: 14-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Paul L.
; REGISTRATION NUMBER: 35,695
; REFERENCE/DOCKET NUMBER: MSP-027
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 7:

```

```
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1004
; US-08-293-300A-7

Query Match      1.3%  Score 57.4; DB 3; Length 1004;
Best Local Similarity 60.6%; Pred. No. 3.5e-07;
Matches 94; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 3785 AATGTGTGCACTGGCATTGCTTTCTGGATGAAGTATGATGCTTTGGCATATGAGT 3844
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      832 AAGTGGCAACACTCCGGGTGCTCTCTGGATCAAGTATGATGCTTTGGCATATGAGT 3844
QY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      3845 TGGCATGTGACAGTGGACCTTTCCAGTGCAGCATATGCCAGCATCTTTGCCAGGACAAA 3904
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      892 TGGCAGGGCTTCGGTATCTCTTGGATGCAACATGTGCGCTTACCACAGCCAGGACCG 951
QY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      3905 TATGACTTCAACACATATCCAGAGGGGCGCTGCA 3939
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      952 TACGAGTTCTCGTGCACATATACGCGAGGGGAGCA 986

RESULT 5
US-08-711-417C-5
; Sequence 8, Application: US/08711417C
; Patent No. 6228611
; GENERAL INFORMATION:
; APPLICANT: Georgopoulos, Katia A.
; TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE
; NUMBER OF SEQUENCES: 202
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/711,417C
; FILING DATE: 05-Sep-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/238,212
; FILING DATE: 02-MAY-1994
; APPLICATION NUMBER: 08/121,438
; FILING DATE: 14-SEP-1993
; APPLICATION NUMBER: 07/946,233
; FILING DATE: 14-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis P.
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10287/007001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
```

```
;
; NAME/KEY: Coding Sequence
; LOCATION: 1..1002
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-08-711-417C-8

Query Match      1.3%  Score 57.4; DB 3; Length 1004;
Best Local Similarity 60.6%; Pred. No. 3.5e-07;
Matches 94; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 3785 AATGTGTGCACTGGCATTGCTTTCTGGATGAAGTATGATGCTTTGGCATATGAGT 3844
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      832 AAGTGGCAACACTCCGGGTGCTCTCTGGATCAAGTATGATGCTTTGGCATATGAGT 3844
QY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      3845 TGGCATGTGACAGTGGACCTTTCCAGTGCAGCATATGCCAGCATCTTTGCCAGGACAAA 3904
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      892 TGGCAGGGCTTCGGTATCTCTTGGATGCAACATGTGCGGTACCACAGCCAGGACCG 951
QY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      3905 TATGACTTCAACACATATCCAGAGGGGCGCTGCA 3939
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      952 TACGAGTTCTCGTGCACATATACGCGAGGGGAGCA 986

RESULT 6
US-08-733-622C-21
; Sequence 21, Application: US/08733622C
; Patent No. 6528634
; GENERAL INFORMATION:
; APPLICANT: Georgopoulos, Katia
; APPLICANT: Morgan, Bruce A.
; TITLE OF INVENTION: AIOLOS GENE
; FILE REFERENCE: 10287-330001
; CURRENT APPLICATION NUMBER: US/08/733,622C
; CURRENT FILING DATE: 1996-10-17
; PRIOR APPLICATION NUMBER: US 60/017,546
; PRIOR FILING DATE: 1996-05-14
; PRIOR APPLICATION NUMBER: US 60/005,529
; PRIOR FILING DATE: 1995-10-18
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1004
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 11...(1002)
US-08-733-622C-21

Query Match      1.3%  Score 57.4; DB 4; Length 1004;
Best Local Similarity 60.6%; Pred. No. 3.5e-07;
Matches 94; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 3785 AATGTGTGCACTGGCATTGCTTTCTGGATGAAGTATGATGCTTTGGCATATGAGT 3844
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      832 AAGTGGCAACACTCCGGGTGCTCTCTGGATCAAGTATGATGCTTTGGCATATGAGT 3844
QY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      3845 TGGCATGTGACAGTGGACCTTTCCAGTGCAGCATATGCCAGCATCTTTGCCAGGACAAA 3904
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      892 TGGCAGGGCTTCGGTATCTCTTGGATGCAACATGTGCGGTACCACAGCCAGGACCG 951
QY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      3905 TATGACTTCAACACATATCCAGAGGGGCGCTGCA 3939
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      952 TACGAGTTCTCGTGCACATATACGCGAGGGGAGCA 986

RESULT 7
PCT-US95-03345-7
; Sequence 7, Application: PCT/US9503345
; GENERAL INFORMATION:
; APPLICANT: Georgopoulos, Katia A.
; TITLE OF INVENTION: IKAROS TRANSGENIC CELLS AND ANIMALS
; NUMBER OF SEQUENCES: 26
```

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: LAHIVE & COCKFIELD  
 STREET: 60 STATE STREET, Suite 510  
 CITY: BOSTON  
 STATE: MASSACHUSETTS  
 COUNTRY: USA  
 ZIP: 02109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: ASCII (text)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US95/09345  
 FILING DATE:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/283,300  
 FILING DATE: 29-JULY-94  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/238,212  
 FILING DATE: 02-MAY-94  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/121,438  
 FILING DATE: 14-SEP-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/946,233  
 FILING DATE: 14-SEP-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Myers, Paul L.  
 REGISTRATION NUMBER: 35,695  
 REFERENCE/DOCKET NUMBER: MGP-027PC  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617)227-7400  
 TELEFAX: (617)227-5941  
 INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1004 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 FEATURES:  
 NAME/KEY: CDS  
 LOCATION: 1..1004  
 PCT-US95-39345-7

Query Match 1.3%; Score 57.4; DB 5; Length 1004;  
 Best Local Similarity 60.6%; Pred. No. 3 5e-07;  
 Matches 94; Conservative 0; Mismatches 61; Indels 0; Gaps 2;  
 QY 3785 AATGTGTGCACGTGGCATTGCTTTCTGGATGAAGTGAATGATGCTTTGTCATATGAGT 3844  
 DB 832 AATGGGACACTGCCGGGTGCTCTTCTGGATGACGTCATGTACACCATCCACATGGGC 891  
 QY 3845 TGGCATGTGCACGTGGCATTGCTTTCCAGTGCAGCATATGCCAGATCTTTGACACGACAAA 3904  
 DB 892 TGCACGGCTTCGTGATCTTTTGATGCAACATGTGCGGTACACAGCAGGACCGG 951  
 QY 3905 TATGACTTCACACATATCCAGAGGGGCTTCCA 3939  
 DB 952 TAGCAGTTCGTGCGACATATCCAGAGGGGACCA 986

RESULT 8  
 US-08-465-590-3  
 Sequence 3, Application US/08465590  
 Patent No. 5824770  
 GENERAL INFORMATION:  
 APPLICANT: Georgopoulos, Katia A.  
 TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE  
 NUMBER OF SEQUENCES: 164  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: LAHIVE & COCKFIELD

STREET: 60 STATE STREET, Suite 510  
 CITY: BOSTON  
 STATE: MASSACHUSETTS  
 COUNTRY: USA  
 ZIP: 02109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: ASCII (text)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/465,590  
 FILING DATE: 05-JUN-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/238,212  
 FILING DATE: 02-MAY-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/121,438  
 FILING DATE: 14-SEP-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/946,233  
 FILING DATE: 14-SEP-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Myers, Paul L.  
 REGISTRATION NUMBER: 35,695  
 REFERENCE/DOCKET NUMBER: MGP-026C2DV  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617)227-7400  
 TELEFAX: (617)227-5941  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1386 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 FEATURES:  
 NAME/KEY: CDS  
 LOCATION: 1..1386  
 US-08-465-590-3

Query Match 1.3%; Score 57.4; DB 1; Length 1386;  
 Best Local Similarity 60.6%; Pred. No. 4 5e-07;  
 Matches 94; Conservative 0; Mismatches 61; Indels 3; Gaps 0;  
 QY 3785 AATGTGTGCACGTGGCATTGCTTTCTGGATGAAGTGAATGCTTTGTCATATGAGT 3844  
 DB 1213 AATGGGACACTGCCGGGTGCTCTTCTGGATGACGTCATGTACACCATCCACATGGGC 1272  
 QY 3845 TGGCATGTGCACGTGGCATTGCTTTCCAGTGCAGCATATGCCAGATCTTTGACACGACAAA 3904  
 DB 1273 TGCACGGCTTCGTGATCTTTTGATGCAACATGTGCGGTACACAGCAGGACCGG 1332  
 QY 3905 TATGACTTCACACATATCCAGAGGGGCTTCCA 3939  
 DB 1333 TAGCAGTTCGTGCGACATATCCAGAGGGGACCA 1367

RESULT 9  
 US-08-283-300A-2  
 Sequence 2, Application US/08283300A  
 Patent No. 6172278  
 GENERAL INFORMATION:  
 APPLICANT: Georgopoulos, Katia A.  
 TITLE OF INVENTION: IKAROS TRANSGENIC CELLS AND ANIMALS  
 NUMBER OF SEQUENCES: 26  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: LAHIVE & COCKFIELD  
 STREET: 60 STATE STREET, Suite 510  
 CITY: BOSTON  
 STATE: MASSACHUSETTS  
 COUNTRY: USA  
 ZIP: 02109

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII (text)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/283,300A  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/238,212  
FILING DATE: 02-MAY-94  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/121,438  
FILING DATE: 14-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/946,233  
FILING DATE: 14-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Myers, Paul L.  
REGISTRATION NUMBER: 35,695  
REFERENCE/DOCKET NUMBER: MGP-027  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1386 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1386  
US-08-283-300A-2

Query Match 1.3%; Score 57.4; DB 3; Length 1386;  
Best Local Similarity 60.6%; Pred. No. 4.5e-07;  
Matches 94; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 3785 AATGTGTGCACCTGGCATGTCTTTCTGGATGAAGTGATGTATGCTTTGCATATAGCT 3844  
Db 1213 AAGTGGCAACACTGCCGGGTGCTCTTCTGGATCAGCTCATGTACACCATCCACATGGGC 1272

Qy 3845 TGCATGTTGACAGTGAGTACCTTTCCATGCGAGCATATGCCAGCATTTTGCACGGACAAA 3904  
Db 1273 TGCACGGCTTCCTGTATCTCTTTTGTGTCAACATGTGCGGTACACACAGCCAGGACCG 1332

Qy 3905 TATGACTTCACACACATATCCAGAGGGGCTGCA 3939  
Db 1333 TACGAGTTCTGTCGCACATAACCGAGGGAGCA 1367

RESULT 10  
US-08-711-417C-3  
Sequence 3, Application US/08/711417C  
Patent No. 6228611  
GENERAL INFORMATION:  
APPLICANT: Georgopoulos, Katia A.  
TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE  
NUMBER OF INVENTENCES: 202  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FastSeq for Windows Version 2.0b

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/711,417C  
FILING DATE: 05-SEP-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/238,212  
FILING DATE: 02-MAY-1994  
APPLICATION NUMBER: 08/121,438  
FILING DATE: 14-SEP-1993  
APPLICATION NUMBER: 07/946,233  
FILING DATE: 14-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Myers, Louis P.  
REGISTRATION NUMBER: 35,965  
REFERENCE/DOCKET NUMBER: 10287/007001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1386 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 1..1383  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-08-711-417C-3

Query Match 1.3%; Score 57.4; DB 3; Length 1386;  
Best Local Similarity 60.6%; Pred. No. 4.5e-07;  
Matches 94; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 3785 AATGTGTGCACCTGGCATGTCTTTCTGGATGAAGTGATGTATGCTTTGCATATAGCT 3844  
Db 1213 AAGTGGCAACACTGCCGGGTGCTCTTCTGGATCAGCTCATGTACACCATCCACATGGGC 1272

Qy 3845 TGCATGTTGACAGTGAGTACCTTTCCATGCGAGCATATGCCAGCATTTTGCACGGACAAA 3904  
Db 1273 TGCACGGCTTCCTGTATCTCTTTTGTGTCAACATGTGCGGTACACACAGCCAGGACCG 1332

Qy 3905 TATGACTTCACACACATATCCAGAGGGGCTGCA 3939  
Db 1333 TACGAGTTCTGTCGCACATAACCGAGGGAGCA 1367

RESULT 11  
US-08-733-622C-16  
Sequence 16, Application US/08/733622C  
Patent No. 6528634  
GENERAL INFORMATION:  
APPLICANT: Georgopoulos, Katia  
APPLICANT: Morgan, Bruce A.  
TITLE OF INVENTION: AICLOS GENE  
FILE REFERENCE: 0287-03000  
CURRENT APPLICATION NUMBER: US/08/733,622C  
CURRENT FILING DATE: 1996-10-17  
PRIOR APPLICATION NUMBER: US 60/017,646  
PRIOR FILING DATE: 1996-05-14  
PRIOR APPLICATION NUMBER: US 60/005,529  
PRIOR FILING DATE: 1995-10-18  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 16  
LENGTH: 1386  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(1383)  
US-08-733-622C-16

Query Match 1.3% Score 57.4; DB 4; Length 1386;  
Best Local Similarity 60.6%; Pred. No. 4.5e-07;  
Matches 94; Conservative 0; Mismatches 61; Indels 0; Gaps 0;  
QY 3785 AATGTGTGCACTGGGCACTTCTTCTGGATGAAGTATGCTTTTGCATATGAGT 3844  
DB 1213 AAGTGGGAACACTGGCGGTGCTCTTCTGGATGACGTCATGTACACCATCCACATGGGC 1272  
QY 3845 TCCCATGGTGCAGTGGGACCTTTCAGTGCACATATCCAGAGGCGCTGCA 3904  
DB 1273 TCCACAGGCTTCGGTGCATCTTTGAGTGCACATGTGCGGTACACAGCGGAGCGG 1332  
QY 3905 TATGACTTCCACACACATATCCAGAGGCGCTGCA 3939  
DB 1333 TAGGACTTCTGTCGCACATACAGCGGAGCGGAGCA 1367

RESULT 12  
PCT-US95-09345-2  
Sequence 165, Application PC/TUS9509345  
GENERAL INFORMATION:  
APPLICANT: Georgopoulos, Katia A.  
TITLE OF INVENTION: IKAROS TRANSGENIC CELLS AND ANIMALS  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSES: LAHIVE & COCKFIELD  
STREET: 60 STATE STREET, Suite 510  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII (text)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/09345  
FILING DATE:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/263,300  
FILING DATE: 29-JULY-94  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/238,212  
FILING DATE: 02-MAY-94  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/121,438  
FILING DATE: 14-SEP-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/946,233  
FILING DATE: 14-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Myers, Paul J.  
REGISTRATION NUMBER: 35,695  
REFERENCE/DOCKET NUMBER: MGP-0272C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1386 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1386  
PCT-US95-09345-2

Query Match 1.3% Score 57.4; DB 5; Length 1386;  
Best Local Similarity 60.6%; Pred. No. 4.5e-07;

Matches 94; Conservative 0; Mismatches 61; Indels 0; Gaps 0;  
QY 3785 AATGTGTGCACTGGGCACTTCTTCTGGATGAAGTATGCTTTTGCATATGAGT 3844  
DB 1213 AAGTGGGAACACTGGCGGTGCTCTTCTGGATGACGTCATGTACACCATCCACATGGGC 1272  
QY 3845 TCCCATGGTGCAGTGGGACCTTTCAGTGCACATATCCAGAGGCGCTGCA 3904  
DB 1273 TCCACAGGCTTCGGTGCATCTTTGAGTGCACATGTGCGGTACACAGCGGAGCGG 1332  
QY 3905 TATGACTTCCACACACATATCCAGAGGCGCTGCA 3939  
DB 1333 TAGGACTTCTGTCGCACATACAGCGGAGCGGAGCA 1367

RESULT 13  
US-08-711-417C-165  
Sequence 165, Application US/08711417C  
Patent No. 6228611  
GENERAL INFORMATION:  
APPLICANT: Georgopoulos, Katia A.  
TITLE OF INVENTION: IKAROS: A T CELL PATHWAY REGULATORY GENE  
NUMBER OF SEQUENCES: 202  
CORRESPONDENCE ADDRESS:  
ADDRESSES: Fisk & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/711,417C  
FILING DATE: 05-SEP-1996  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 08/238,212  
FILING DATE: 02-MAY-1994  
APPLICATION NUMBER: 08/121,438  
FILING DATE: 14-SEP-1993  
APPLICATION NUMBER: 07/946,233  
FILING DATE: 14-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Myers, Louis P.  
REGISTRATION NUMBER: 35,965  
REFERENCE/DOCKET NUMBER: 20297/007601  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 165:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1551 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 1..1548  
SEQUENCE DESCRIPTION: SEQ ID NO: 165:  
US-08-711-417C-165

Query Match 1.3% Score 57.4; DB 3; Length 1551;  
Best Local Similarity 60.6%; Pred. No. 4.9e-07;  
Matches 94; Conservative 0; Mismatches 61; Indels 0; Gaps 0;  
QY 3785 AATGTGTGCACTGGGCACTTCTTCTGGATGAAGTATGCTTTTGCATATGAGT 3844  
DB 1378 AAGTGGGAACACTGGCGGTGCTCTTCTGGATGACGTCATGTACACCATCCACATGGGC 1437



3845 TGCCATGGTGACAGTGGAGCTTTCCAGTCGAGCATATGCCAGCATCTTTCCACGGACAAA 3904  
1438 TGCCACGGCTTCGTGATCTTTTGAGTCAACATGTGCGGTACCACACCCAGGACCG 1497

QY  
Db  
QY  
Db

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	1.34;	Score 57.4;	DB 5;	Length 1611;
Best Local Similarity	60.6%;	Pred. No. Se-07;		
Matches 94;	Conservative	0; Mismatches 61;	Indels 0;	Gaps 0;
Qy	3785	AAATGTGTGCACCTGTGGCATTTCTTCTGTGATGAAGTGATGTATGCTTTGCATATCAGT	3844	
Db	1435	AAGTGCACAACACTCCGGTGCTCTTCTTGATCAGTCTATGTACATCCTCCACATGGCC	1494	
Qy	3845	TGCCATGTGTGACAGTGGAGCCTTTCCAGTGCAGCATATGCCAGCATCTTTTGCACGGACAA	3904	
Db	1495	TGCCACGGCTTCGTGATCCTTTTGTAGTGCACCATGTGGCGCTACCACGCCAGACCGG	1554	
Qy	3905	TATGACTTTCAACAACATATCCAGAGGGGCGCTGCA	3939	
Db	1555	TACGAGTTCTCGTCGCATAAACCGAGGGGAGCA	1589	

```

RESULT 15
US-09-435-327A-19
; Sequence 19, Application US/09435327A
; Patent No. 653766
; GENERAL INFORMATION:
; APPLICANT: Uckun, Fatih M.
; APPLICANT: Crotty, Mya L.
; TITLE OF INVENTION: IKAROS ISOFORMS AND MUTANTS
; FILE REFERENCE: 12152.35USU1
; CURRENT APPLICATION NUMBER: US/09/435,327A
; CURRENT FILING DATE: 1999-11-05
; PRIOR APPLICATION NUMBER: 60/107,229
; PR-OR FILING DATE: 1998-11-05
; NUMBER OF SEQ ID NOS: 27

```

```

; SOFTWARE: PatenIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 1788
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-435-327A-19

Query Match: 1.3%; Score 57.4; DB 4; Length 1788;
Best Local Similarity 60.6%; Pred. No. 5.4e-07;
Matches 94; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 3785 AAATGTGTGCACGTGTGCGATGTCTTTCTGGATGAAGTGTATGCTTTGGCATATGAGT 3844
1615 AAGTGGGAACACATGCCGGGTGCTCTTCTCGATCAGGTGATGACACATCCACATGGGC 1674
3845 TGCCATGGTGACAGTGGACGCTTTCCAGTGACAGCATATGCCAGCATCTTTTCACGGACAAA 3904
1675 TGCACGGCTTCGGTGATCTTTTGAGTGGAACATGTGCGGCTACACACCCAGGACCGG 1734
3905 TATGACTTCACAAACACATATCCAGAGGGGCTTCCA 3939
1735 TAGCAGTCTTCGTGGCACAATAGCCGAGGGGAGCA 1769
Db

```

Search completed: October 29, 2003, 15:29:16  
Job time : 237 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 29, 2003, 11:49:17 ; Search time 1125 Seconds

(without alignments)

10881.588 Million cell updates/sec

Title: US-09-702-216-1\_COPY\_500\_5000

Perfect score: 4501

Sequence: 1 atgtcttctgtcttctga.....agaaaaatggccctctca 450:

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1811591 seqs, 1359896230 residues

Total number of hits satisfying chosen parameters: 3623182

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

```
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCr_NEW_PJB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US05_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCrUS_PJBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4501	100.0	10011	14	US-10-177-293-471
2	4501	100.0	10102	14	US-10-198-846-10440
3	1060.4	23.6	1062	12	US-10-029-386-25248
4	542.2	12.0	879	14	US-10-198-846-3873
5	510.4	11.3	512	12	US-10-029-386-11541
6	453.4	10.1	510	14	US-10-198-846-3994
7	398.8	8.9	520	14	US-10-198-846-8958
8	389.2	8.6	836	14	US-10-198-846-3330
9	324.4	7.2	427	14	US-10-198-846-5878
10	322.8	7.2	476	11	US-09-918-995-10414
11	275.4	6.1	305	9	US-09-778-320-110
12	275.4	6.1	305	9	US-09-910-689-110
13	275.4	6.1	305	13	US-10-010-742-110
14	210	4.7	810	14	US-10-198-846-7424
15	133.4	3.0	594	14	US-10-198-846-9634
16	119.2	2.6	467	14	US-10-198-846-419

17	99.4	2.2	968	14	US-10-198-846-7398	Sequence 7398, Ap
18	80.6	1.8	166	10	US-09-728-444-1065	Sequence 1065, Ap
19	65.4	1.5	2177	14	US-10-084-817-291	Sequence 291, App
20	65.4	1.5	2610	14	US-10-177-293-173	Sequence 173, App
21	65	1.4	569	14	US-10-198-846-11151	Sequence 11151, A
22	65	1.4	1026	14	US-10-198-846-10906	Sequence 10906, A
23	65	1.4	1475	10	US-09-964-824A-293	Sequence 293, App
24	65	1.4	2025	15	US-10-081-327-46	Sequence 46, Appl
25	65	1.4	2365	12	US-10-097-926A-78	Sequence 78, Appl
26	65	1.4	2673	14	US-10-177-293-178	Sequence 178, App
27	65	1.4	2673	14	US-10-177-293-177	Sequence 177, App
28	65	1.4	2709	14	US-10-084-817-331	Sequence 331, App
29	65	1.4	3028	14	US-10-177-293-175	Sequence 175, App
30	65	1.4	3118	14	US-10-198-846-11426	Sequence 11426, A
31	64.6	1.4	514	10	US-09-764-864-196	Sequence 196, App
32	64.6	1.4	523	10	US-09-764-864-1623	Sequence 1623, Ap
33	64.6	1.4	523	10	US-09-764-864-1624	Sequence 1624, Ap
34	64.6	1.4	523	10	US-09-764-864-1625	Sequence 1625, Ap
35	63.6	1.4	674	14	US-10-198-846-7898	Sequence 7898, Ap
36	63.4	1.4	1326	10	US-09-749-728B-12	Sequence 12, Appl
37	63.4	1.4	3411	14	US-10-198-846-9872	Sequence 9872, Ap
38	62.8	1.4	3671778	12	US-10-312-841-2	Sequence 2, Appl
39	62	1.4	682	14	US-10-198-846-2403	Sequence 2403, Ap
40	57.4	1.3	1024	10	US-09-755-830-7	Sequence 7, Appl
41	57.4	1.3	1386	10	US-09-755-830-2	Sequence 2, Appl
42	56.6	1.3	2453	14	US-10-198-846-13080	Sequence 13080, A
43	55.8	1.2	1498	12	US-10-007-926A-4	Sequence 4, Appl
44	55.6	1.2	856	14	US-10-198-846-4034	Sequence 4034, Ap
45	53.9	1.2	1647	14	US-10-106-698-1074	Sequence 1074, Ap

#### ALIGNMENTS

#### RESULT 1

```
US-10-177-293-471
; Sequence 471, Application: US/10177293
; Publication No. US2003012428A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Ghatt, Karen
; APPLICANT: Zhao, Xumel
; APPLICANT: Gannavarpu, Manjula
; APPLICANT: Kamatkar, Shubhang
; APPLICANT: Vertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Puzstai, LaJos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
```

PRIOR FILING DATE: 2002-05-14  
; NUMBER OF SEQ ID NOS: 506  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 471  
; LENGTH: 10011  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-177-293-471

Query Match 100.0%; Score 4501; DB 14; length 10011;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 4501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	ATGTCCTCTGTCCTCTGAGCGTCCTCTAGACAAATTCGATTAGCCTCTGCTAGTTGA	60
Db	500	ATGTCCTCTGTCCTCTGAGCGTCCTCTAGACAAATTCGATTAGCCTCTGCTAGTTGA	559
Qy	61	CTAATAGAAATTAATATGTAATAAGCACTCTAAAGCCACATGCCCTTATGAAGTCAATGC	120
Db	560	CTAATAGAAATTAATATGTAATAAGCACTCTAAAGCCACATGCCCTTATGAAGTCAATGC	619
Qy	121	TGGGTATGATTTTACAATATGTTCCGGAAGAAACCCCCCTCTGAGAACCTTSCAAG	180
Db	620	TGGGTATGATTTTACAATATGTTCCGGAAGAAACCCCCCTCTGAGAACCTTSCAAG	679
Qy	181	TGAAGGCGAGGCGCAGATCTCTGGAGCCTATAGGTACAGAAGCAAGGTATCTGGAAGAA	240
Db	680	TGAAGGCGAGGCGCAGATCTCTGGAGCCTATAGGTACAGAAGCAAGGTATCTGGAAGAA	739
Qy	241	CAAGAATTTCTCGAGATCAGATGTGCAAAAATACCGATCAGAGTGAATCTCGAAGCT	300
Db	740	CAAGAATTTCTCGAGATCAGATGTGCAAAAATACCGATCAGAGTGAATCTCGAAGCT	799
Qy	301	AAATCATAGAGGAGCAATACCTTGCTGATGTTCTAGATCCATCTTCTAGCAGTAAAGGA	360
Db	800	AAATCATAGAGGAGCAATACCTTGCTGATGTTCTAGATCCATCTTCTAGCAGTAAAGGA	859
Qy	361	CTTGAAGAAGCCAGTCTGAGTGAGAGGCTGGCTTCAATATTAAGAAAGCCAGTAAAGG	420
Db	860	CTTGAAGAAGCCAGTCTGAGTGAGAGGCTGGCTTCAATATTAAGAAAGCCAGTAAAGG	919
Qy	421	AGGAACCTTCTGCTTCCGATGATGAGGTGACAGACAGAAATATGTTGGCTTCTC	480
Db	920	AGGAACCTTCTGCTTCCGATGATGAGGTGACAGACAGAAATATGTTGGCTTCTC	979
Qy	481	ATTTCCAGCTGCTGGGGAGTCTGTGAGCCCTTCAAGTCTCCGAAAGACAGAGGAGGA	540
Db	980	ATTTCCAGCTGCTGGGGAGTCTGTGAGCCCTTCAAGTCTCCGAAAGACAGAGGAGGA	1039
Qy	541	TGACCTCTCAAGATATGGCTTGACCCCTCAGGGAATCTCATGTGAGACAAAGGAGATCA	600
Db	1040	TGACCTCTCAAGATATGGCTTGACCCCTCAGGGAATCTCATGTGAGACAAAGGAGATCA	1099
Qy	601	GAGGATGTCACCAAGGCTACAGAGGAACAGGCAAGCAGAGTGGTCAGGCGCAATTC	660
Db	1100	GAGGATGTCACCAAGGCTACAGAGGAACAGGCAAGCAGAGTGGTCAGGCGCAATTC	1159
Qy	661	TCAAGGTTTGAGCCAGTTTCAGTGGCTCAAAAACCCCAAGTGGCTTCAAGTGGGG	720
Db	1160	TCAAGGTTTGAGCCAGTTTCAGTGGCTCAAAAACCCCAAGTGGCTTCAAGTGGGG	1219
Qy	721	TGTAAGACTGAATTAATCGAAAATCTTACTGTGTAATGACACCCAGACCCGGGACC	780
Db	1220	TGTAAGACTGAATTAATCGAAAATCTTACTGTGTAATGACACCCAGACCCGGGACC	1279
Qy	781	TCGTGCTCCAGAGCTTCAAGACTTTAAATGCAATATCTGTGATATGTTACTACGCAAA	840
Db	1280	TCGTGCTCCAGAGCTTCAAGACTTTAAATGCAATATCTGTGATATGTTACTACGCAAA	1339
Qy	841	CGACCCCAAGACTCTGATTAAAGCACTTCGAAAGTATCACTTAGGACTGCAATACCCAC	900
Db	1340	CGACCCCAAGACTCTGATTAAAGCACTTCGAAAGTATCACTTAGGACTGCAATACCCAC	1399

Qy	901	CAGCAAGATGCTCAGCTGGACACCAAAATCTTGGCCCTTCATAAATGCTGTCAGTTCAG	960
Db	1400	CAGCAAGATGCTCAGCTGGACACCAAAATCTTGGCCCTTCATAAATGCTGTCAGTTCAG	1459
Qy	961	CCATTCCAAAGACTTCCAGAAAGTCAACCGTCTCTGTTTCTGGTGTGCTGAGAGACAT	1020
Db	1460	CCATTCCAAAGACTTCCAGAAAGTCAACCGTCTCTGTTTCTGGTGTGCTGAGAGACAT	1519
Qy	1021	CAATCTTCAAGGCTGTTTCTAATAAGGACCTATGATGTCAGAGTCACTTCAAGTGG	1080
Db	1520	CAATCTTCAAGGCTGTTTCTAATAAGGACCTATGATGTCAGAGTCACTTCAAGTGG	1579
Qy	1081	AACATTCATTCGGCATTCGAGGAAACACACAGATTCGCAAGGGAACACCAAGTATTC	1140
Db	1580	AACATTCATTCGGCATTCGAGGAAACACACAGATTCGCAAGGGAACACCAAGTATTC	1639
Qy	1141	CTGTAATCTGSCAATTCCTATATGSGCAACTCATCCACGAAATAGAACACATTT	1200
Db	1640	CTGTAATCTGSCAATTCCTATATGSGCAACTCATCCACGAAATAGAACACATTT	1699
Qy	1201	TCCTCAGACTCACCAAAACAAATATAAGGCTTCTCCCTCTCTGAGTTCGCAAAAC	1260
Db	1700	TCCTCAGACTCACCAAAACAAATATAAGGCTTCTCCCTCTCTGAGTTCGCAAAAC	1759
Qy	1261	TTCCAGAGAAAACTCTAACAGTCCATCCCTGCACTTCAATCCAGTGAATCTGAGACT	1320
Db	1760	TTCCAGAGAAAACTCTAACAGTCCATCCCTGCACTTCAATCCAGTGAATCTGAGACT	1819
Qy	1321	GGGAATGGCAGAGCAAGATTAACAGTCAAGAGGAGGATGACACTCTGTTGGTACTC	1380
Db	1820	GGGAATGGCAGAGCAAGATTAACAGTCAAGAGGAGGATGACACTCTGTTGGTACTC	1879
Qy	1381	AGTSCCACTAAAGCCCTCGATTCCTCTAGACAAATGCTACAGAGCCCACTTACTA	1440
Db	1880	AGTSCCACTAAAGCCCTCGATTCCTCTAGACAAATGCTACAGAGCCCACTTACTA	1939
Qy	1441	CTGGTGTAAATTTGTAGTTTACGTGTGAGTCAATCTAGTCACTTAACTGTAGAAC	1500
Db	1940	CTGGTGTAAATTTGTAGTTTACGTGTGAGTCAATCTAGTCACTTAACTGTAGAAC	1999
Qy	1501	TTATGSCAGAGCAGCAGGAGCTGAGTGGGGCTTAAATCCAGATTAATGATAA	1560
Db	2000	TTATGSCAGAGCAGCAGGAGCTGAGTGGGGCTTAAATCCAGATTAATGATAA	2059
Qy	1561	GTTTTCCAGGCGCTCTGTCTATTAATCAAGATTAATCTAGCCCAAAAGTTCAGAGGAG	1620
Db	2060	GTTTTCCAGGCGCTCTGTCTATTAATCAAGATTAATCTAGCCCAAAAGTTCAGAGGAG	2119
Qy	1621	AATGACCAAGACAGACAGCTCGAGTGGGGCTTAAAGAGAGGACTTCTCCAGAGGG	1680
Db	2120	AATGACCAAGACAGACAGCTCGAGTGGGGCTTAAAGAGAGGACTTCTCCAGAGGG	2179
Qy	1681	AGCCAGGATTAATATGTAAGGCTAATTTGTCAGTCTGTCGACTTCGATATTCGAA	1740
Db	2180	AGCCAGGATTAATATGTAAGGCTAATTTGTCAGTCTGTCGACTTCGATATTCGAA	2239
Qy	1741	AAGCCATGGCCCTGATTAATTTAGTGGGGCTTCTGCTGATTAATCAAGAGTCCA	1800
Db	2240	AAGCCATGGCCCTGATTAATTTAGTGGGGCTTCTGCTGATTAATCAAGAGTCCA	2299
Qy	1801	TACATTCACAGTCTACATTAACACTGTCCTCTGTCAGAGAGGACTTTTCAGGCC	1860
Db	2300	TACATTCACAGTCTACATTAACACTGTCCTCTGTCAGAGAGGACTTTTCAGGCC	2359
Qy	1861	AGAAAAGCAGCTTTGAGAAATTAATTTGCTGTTGTAAGAAAAAGTAAATTTGCCA	1920
Db	2360	AGAAAAGCAGCTTTGAGAAATTAATTTGCTGTTGTAAGAAAAAGTAAATTTGCCA	2419
Qy	1921	CTGTCACACTCTGTTCTGCACTTGTCTCTGGGGCGGCTGGAAGCTCGGAGTCAACA	1980
Db	2420	CTGTCACACTCTGTTCTGCACTTGTCTCTGGGGCGGCTGGAAGCTCGGAGTCAACA	2479
Qy	1981	TCAGTGCCATCAGTGTTCATTCACCAACCCCTGAGGTAGATGTACTCTCTTCTACTATGA	2040

2480	DB		TCAGTGC	CACTAGTGTTCATCTACCA	CCCGTACGATGTATCTCCTCTTTTCACTATGA	2539
2041	QY		AACTGTGC	ATGAGTCCCAAGCATCGCATGTCAAA	CAAGAAGCAAAATCACTCTGCAAGGATC	2100
2540	DB		AACTGTGC	ATGAGTCCCAAGCATCGCATGTCAAA	CAAGAAGCAAAATCACTCTGCAAGGATC	2599
2101	QY		GGATGGC	AGCAGCTGTGTCAAGGAAAGCAAA	CACTCATGTACCAATGTGATTTTAT	2160
2600	DB		GGATGGC	AGCAGCTGTGTCAAGGAAAGCAAA	CACTCATGTACCAATGTGATTTTAT	2659
2161	QY		TACCAAGT	CGAAGAGAGATTTCCG	CACACTACAGAGAGCACACAGTGTCTACAAATG	2220
2660	DB		TACCAAGT	CGAAGAGAGATTTCCG	CACACTACAGAGAGCACACAGTGTCTACAAATG	2719
2221	QY		CGCTCAGT	GCAGTHTTACAGCTCCGATAC	TCTCAGTCTCTGAGACCTTCCAACTGT	2280
2720	DB		CGCTCAGT	GCAGTHTTACAGCTCCGATAC	TCTCAGTCTCTGAGACCTTCCAACTGT	2779
2281	QY		TCACTGCC	AGACAGACATCACTACAGCC	ACCGGAGAGGAGCGGTCTATGCCATAC	2340
2780	DB		TCACTGCC	AGAAACAGACATCACTACAGCC	ACCGGAGAGGAGCGGTCTATGCCATAC	2839
2341	QY		CACCAT	CAAAAGAGAGCCAAAATTG	ACTTTCAGGGGTCTACAAATCTGCTAACTCCAGACTC	2400
2840	DB		CACCAT	CAAAAGAGAGCCAAAATTG	ACTTTCAGGGGTCTACAAATCTGCTAACTCCAGACTC	2899
2401	QY		TAAAAATGG	GAGAGCGAGTTTCTCAGAGATGTGTG	TAAAGAGAGAAAGCTGGAAGAGAGGA	2460
2900	DB		TAAAAATGG	GAGAGCGAGTTTCTCAGAGATGTGTG	TAAAGAGAGAAAGCTGGAAGAGAGGA	2959
2461	QY		CGGGCTCAA	AGAGAAAGTTTGGACCCGAGAGTTC	CAGTGATGACCTTCGCAATGTGACTTG	2520
2960	DB		CGGGCTCAA	AGAGAAAGTTTGGACCCGAGAGTTC	CAGTGATGACCTTCGCAATGTGACTTG	3019
2521	QY		GAGAGGGC	CAGACATCTCTCGGGGGAGTCCGT	CATACACCCAAAGCAAGCTGGGGTGTCT	2580
3020	DB		GAGAGGGC	CAGACATCTCTCGGGGGAGTCCGT	CATACACCCAAAGCAAGCTGGGGTGTCT	3079
2581	QY		GAGCGCTGT	CTGGACCCAGAGCAGACAAAGACTCT	TAAGGATAGTCCCAATGTGGA	2640
3080	DB		GAGCGCTGT	CTGGACCCAGAGCAGACAAAGACTCT	TAAGGATAGTCCCAATGTGGA	3139
2641	QY		GGCGCGCC	ATCTGGCGGACCTATTATTATGGCT	TGGCTGTGGAAACCAAGGGATCTCTGCA	2700
3140	DB		GGCGCGCC	ATCTGGCGGACCTATTATTATGGCT	TGGCTGTGGAAACCAAGGGATCTCTGCA	3199
2701	QY		GGGGGCGC	AGCTGGCGAGAGAGTCTGGGGCCCT	CCCCACAGCATCTCTCATCTGGG	2760
3200	DB		GGGGGCGC	AGCTGGCGAGAGAGTCTGGGGCCCT	CCCCACAGCATCTCTCATCTGGG	3259
2761	QY		AGAAAAC	AAAGTGAATCCCAAGTCCCTGT	TTACGGAGCGGTAGAGGCTCCGGTGT	2820
3260	DB		AGAAAAC	AAAGTGAATGAATCCCAAGTCCCTGT	TTACGGAGCGGTAGAGGCTCCGGTGT	3319
2821	QY		TTTTTTG	GCCAATTCCTCTGACCAAGACCTCT	CTCTCTGCGGAAGAAATGCAATGGCGG	2880
3320	DB		TTTTTTG	GCCAATTCCTCTGACCAAGACCTCT	CTCTCTGCGGAAGAAATGCAATGGCGG	3379
2881	QY		ATATGTAT	GCAAACCGCTGTGCCTCTAC	CAGAGCTTCACTCGACTCCCAAGGCTTTAAA	2940
3380	DB		ATATGTAT	GCAAACCGCTGTGCCTCTAC	CAGAGCTTCACTCGACTCCCAAGGCTTTAAA	3439
2941	QY		CATCATTTAA	CAAAACAAACGGTGAGCAGATTAT	TATTAAGAGGAGAAACAAGAAGCGCTTAA	3000
3440	DB		CATCATTTAA	CAAAACAAACGGTGAGCAGATTAT	TATTAAGAGGAGAAACAAGAAGCGCTTAA	3499
3001	QY		CCCAGAGC	CACTTCCAGGCTGAGCAGCTCA	CAAAACACAGAGGGCGCAGCAATGAGGAGCA	3060
3500	DB		CCCAGAGC	CACTTCCAGGCTGAGCAGCTCA	CAAAACACAGAGGGCGCAGCAATGAGGAGCA	3559
3061	QY		AGTCAAT	TGGAAAGCCCGTTAGAGAGGAGGT	CAGAAAGATCATCTTAATCTGAAAGTCAACAGAG	3120

D5	3560	AGTCAATGGAAGCCCGTTAGAGAGGAGGTCAGAAAGTATCATCTAACTGAAAGTCACCAGAG	3619
QY	3121	AGAAAATCCACTCCCAGCCCTAAAGTAAATACGAAGGCCCGAGGGTTCATTGACTAAAAAGCCA	3180
DB	3620	AGAAAATCCACTCCCAGCCCTAAAGTAAATACGAAGGCCCGAGGGTTCATTGACTAAAAAGCCA	3679
QY	3181	TTCTGCTCAGCAGCCAGTCCCTGGTCAGCCNAACTCTGGATATTACAAAAGGATGCAACC	3240
DB	3680	TTCTGCTCAGCAGCCAGTCCCTGGTCAGCCNAACTCTGGATATTACAAAAGGATGCAACC	3739
QY	3241	TTTGCACTATCAGATAAAAGTCCCTCAGGAAAGTACTCGAGATCCAGGAATAAGTTCATC	3300
DB	3740	TTTGCACTATCAGATAAAAGTCCCTCAGGAAAGTACTCGAGATCCAGGAATAAGTTCATC	3799
QY	3301	CGTATCTGAAGGGAGAAAGAGTTCTCGAGAGGAGCAGTCCCTATAGAAAAGTACATAGAGACC	3360
DB	3800	CGTATCTGAAGGGAGAAAGAGTTCTCGAGAGGAGCAGTCCCTATAGAAAAGTACATAGAGACC	3859
QY	3361	TGGGAACACCCAAATATTATTCACACACAGCAGCCCTATTGAAAAGTACCAGTACCCACT	3420
DB	3860	TGGGAACACCCAAATATTATTCACACACAGCAGCCCTATTGAAAAGTACCAGTACCCACT	3919
QY	3421	TTTGTGACTTCCCTTTGTACATTAATGACTTCCAGAGTGAAGCTGATTGGCTGGGTTCTG	3480
DB	3920	TTTGTGACTTCCCTTTGTACATTAATGACTTCCAGAGTGAAGCTGATTGGCTGGGTTCTG	3979
QY	3481	GAGTAAATATAGCTCTCCCTTCTCGGNATCCGCTACTCTTGAGTACACCTGCTGCCT	3540
DB	3980	GAGTAAATATAGCTCTCCCTTCTCGGAAATCCGCACTCTTGAGTACACCTGCTGCCT	4039
QY	3541	ACCAAAATCCCTTGGCAAAACTATGTGCTTATCCCACTTCAATTCGCTCCTCATATTTTC	3600
DB	4040	ACCAAAATCCCTTGGCAAAACTATGTGCTTATCCCACTTCAATTCGCTCCTCATATTTTC	4099
QY	3601	AGCTGTTGGATCAGACAAATGACATTCCTCTAGATTGTGCGATCMAGCATTTCCAGACCTGG	3660
DB	4100	AGCTGTTGGATCAGACAAATGACATTCCTCTAGATTGTGCGATCMAGCATTTCCAGACCTGG	4159
QY	3661	GCCAAATCCGAAAACGGTGCCTTCCGAAGGAGAAAACGAAGGCACCACTAATGTAAAAAATGA	3720
DB	4160	GCCAAATCCGAAAACGGTGCCTTCCGAAGGAGAAAACGAAGGCACCACTAATGTAAAAAATGA	4219
QY	3721	AGTGCCCTTTGAATSTAGTATAAAACAGAGAAAGTTGATAGAAAGTACTCAAGATGAACCTTC	3780
DB	4220	AGTGCCCTTTGAATSTAGTATAAAACAGAGAAAGTTGATAGAAAGTACTCAAGATGAACCTTC	4279
QY	3781	AACAAAATGTGCACTGTGCACTTGCTTCTGATGAAAGTATGATGCTTTGCAAT	3840
DB	4280	AACAAAATGTGCACTGTGCACTTGCTTCTGATGAAAGTATGATGCTTTGCAAT	4339
QY	3841	GAGTTGCCATGGTGACAGTGGACCTTTCCAGTGCAGCATATGCGCAGCATCTTTGCACGGA	3900
DB	4340	GAGTTGCCATGGTGACAGTGGACCTTTCCAGTGCAGCATATGCGCAGCATCTTTGCACGGA	4399
QY	3901	CAAAATATGACTTCAACACATATCCAGAGGGGCTCCATAGGAACAAATGCACAAGTGGGA	3960
DB	4400	CAAAATATGACTTCAACACATATCCAGAGGGGCTCCATAGGAACAAATGCACAAGTGGGA	4459
QY	3961	AAAAAATGGAAAACCTTAAGAGATAAAAACCTTAGCACTTAGCAATTAATAGAAATAGG	4020
DB	4460	AAAAAATGGAAAACCTTAAGAGATAAAAACCTTAGCACTTAGCAATTAATAGAAATAGG	4519
QY	4021	TTTTCTTGATGGGAATTCATAGCTTGTATGCTTTATGAAGCCTATTAAAAAATACT	4080
DB	4520	TTTTCTTGATGGGAATTCATAGCTTGTATGCTTTATGAAGCCTATTAAAAAATACT	4579
QY	4081	TCAATAGAGCCTGCCTTATCCAAACATGAAATTCCTCTTTTGTTATTCCTTTCTTTGATG	4140
DB	4580	TCAATAGAGCCTGCCTTATCCAAACATGAAATTCCTCTTTTGTTATTCCTTTCTTTGATG	4639
QY	4141	AGTAGGTTACCAAGATTAAAAAGTGAGATAAATGGTCAATGAGAAAGAAATGGAAGAATGGT	4200
DB	4640	AGTAGGTTACCAAGATTAAAAAGTGAGATAAATGGTCAATGAGAAAGAAATGGAAGAATGGT	4699



Db 1640 CTGTAAATTCGCAATTTCACTTATATGGCAACTCATCCACGAAATTAGAACACACATTT 1699  
Qy 1201 TCTTCAGACTCACCCAAACAAAATAAAGCTTCTCTCCCTCTCTCTGAGTTGCAAAACC 1260  
Db 1700 TCTTCAGACTCACCCAAACAAAATAAAGCTTCTCTCCCTCTCTCTGAGTTGCAAAACC 1759  
Qy 1261 TTCCAGAGAAAATCTTAACAGTCCATCCCTGACCTTCAATCCAGTGATCTGGAGACTT 1320  
Db 1760 TTCCAGAGAAAATCTTAACAGTCCATCCCTGACCTTCAATCCAGTGATCTGGAGACTT 1819  
Qy 1321 GGGAAAATGGCAGGACAAAGATAACAGTCAAAAGCAGGAGATGACACTCTCTGTTGGTACTC 1380  
Db 1920 GGGAAAATGGCAGGACAAAGATAACAGTCAAAAGCAGGAGATGACACTCTCTGTTGGTACTC 1879  
Qy 1381 AGTGCCCATAAAGCCCTCGATTCTCTAGACAAAATGGTACAGAGCCACCAAGTTACTA 1440  
Db 1880 AGTGCCCATAAAGCCCTCGATTCTCTAGACAAAATGGTACAGAGCCACCAAGTTACTA 1939  
Qy 1441 CTGGTGTAAATTTGTAGTTTCAGCTGTGAGTCATCTAGCTCACTTAAACTCTAGAAC 1500  
Db 1940 CTGGTGTAAATTTGTAGTTTCAGCTGTGAGTCATCTAGCTCACTTAAACTCTAGAAC 1999  
Qy 1501 TTATGGCAAGCAGCAGGACAGTGCAGTCAGCGGCCCTTAATCCAGAGTTAAATGATAA 1560  
Db 2000 TTATGGCAAGCAGCAGGACAGTGCAGTCAGCGGCCCTTAATCCAGAGTTAAATGATAA 2059  
Qy 1561 GCTTTCCAGGGGCTCTGCTAATTAATCAGATGATCTAGCCAAAGTTCCAGAGGAGAC 1620  
Db 2060 GCTTTCCAGGGGCTCTGCTAATTAATCAGATGATCTAGCCAAAGTTCCAGAGGAGAC 2119  
Qy 1621 AATGACCAAGACAGCAAGAGCTCGAGTGGGGCTAAAAAGAGGACTTCTCCAGCAAGG 1680  
Db 2120 AATGACCAAGACAGCAAGAGCTCGAGTGGGGCTAAAAAGAGGACTTCTCCAGCAAGG 2179  
Qy 1681 AGCCGAGGATAATATGGTAACGAGCTAATAATGTGAGTTCTGTGACTTCGGATATCCAA 1740  
Db 2180 AGCCGAGGATAATATGGTAACGAGCTAATAATGTGAGTTCTGTGACTTCGGATATCCAA 2239  
Qy 1741 AAGCCATGGCCCTGATGTAATTTAGTGGGGCCACTTCTCCGTATTATCAACAGTCCA 1800  
Db 2240 AAGCCATGGCCCTGATGTAATTTAGTGGGGCCACTTCTCCGTATTATCAACAGTCCA 2299  
Qy 1801 TAACATTCACAAGTGTACCAATTAACACTGTCCATTTCTGTCACAGAGACTTTGACGCC 1860  
Db 2300 TAACATTCACAAGTGTACCAATTAACACTGTCCATTTCTGTCACAGAGACTTTGACGCC 2359  
Qy 1861 AGAAAGCACCCTGGAGAAATTAATTCGTTTGGTTGTTAGTAAAGTAATTTGTTCCCA 1920  
Db 2360 AGAAAGCACCCTGGAGAAATTAATTCGTTTGGTTGTTAGTAAAGTAATTTGTTCCCA 2419  
Qy 1921 CTGTGCACTCTTGCTTCTGCACTTGTCTCCTGGGGGGCTGGAAGCTCGCGAGTCAAA 1980  
Db 2420 CTGTGCACTCTTGCTTCTGCACTTGTCTCCTGGGGGGCTGGAAGCTCGCGAGTCAAA 2479  
Qy 1981 TCAGTGCCATCAGTGTTCATTCACCAACCCTGAGGTAGATGACTCTCTTTCATGACA 2040  
Db 2480 TCAGTGCCATCAGTGTTCATTCACCAACCCTGAGGTAGATGACTCTCTTTCATGACA 2539  
Qy 2041 AAGTGTGCATGAGTCCCAAGCATCGGATGTCAAAACAAAGAACAAATCACCTGCAAGGATC 2100  
Db 2540 AAGTGTGCATGAGTCCCAAGCATCGGATGTCAAAACAAAGAACAAATCACCTGCAAGGATC 2599  
Qy 2101 GGATGGGAGCAGTCTGTCAAGGAAAGCAAGAACACTCATGTACCAATGTGATTTTC 2160  
Db 2600 GGATGGGAGCAGTCTGTCAAGGAAAGCAAGAACACTCATGTACCAATGTGATTTTC 2659  
Qy 2161 TACCAGTGTGAAGAGAGATTTCCGACACTACAGAGAGCAGCAGCTGCTACAAATG 2220  
Db 2660 TACCAGTGTGAAGAGAGATTTCCGACACTACAGAGAGCAGCAGCTGCTACAAATG 2719  
Qy 2221 CCGTCAGTGCAGTTTACAGTCCGCGATCTCAGTCACTACTGGAGCACTTCAACACTGT 2280  
Db 2720 CCGTCAGTGCAGTTTACAGTCCGCGATCTCAGTCACTACTGGAGCACTTCAACACTGT 2779

Qy 2281 TCACTGCCAGGAACAGGACATCACTACAGCCAAACGGCGAAGAGGAGGTCTATCGCCATATC 2340  
Db 2780 TCACTGCCAGGAACAGGACATCACTACAGCCAAACGGCGAAGAGGAGGTCTATCGCCATATC 2839  
Qy 2341 CACCATCAAGAGAGGCCCAAAATGACITTCAGGTCTACATCTGTAACATCCAGACTC 2400  
Db 2840 CACCATCAAGAGAGGCCCAAAATGACITTCAGGTCTACATCTGTAACATCCAGACTC 2899  
Qy 2401 TAAAAATGGAGAGCCAGTTTCTGAGAGTGTGTGAAGAGAGAGAGCTGGAAGAAAGGA 2460  
Db 2900 TAAAAATGGAGAGCCAGTTTCTGAGAGTGTGTGAAGAGAGAGAGCTGGAAGAAAGGA 2959  
Qy 2461 CCGGCTCAAGAGAAAATTTGGACGAGAGTTTCCAGTGATGACCTTCCGAATGTGACTTG 2520  
Db 2960 CCGGCTCAAGAGAAAATTTGGACGAGAGTTTCCAGTGATGACCTTCCGAATGTGACTTG 3019  
Qy 2521 GAGAGGGGACAGACATCTCGCGGGAGTCCGTTCATACACCCAAAGCAAGCTCGGCTGT 2580  
Db 3020 GAGAGGGGACAGACATCTCGCGGGAGTCCGTTCATACACCCAAAGCAAGCTCGGCTGT 3079  
Qy 2581 GACGCTGTGTGTGGACCCCAAGAGAGAGACAAAGACTCTAAGGGATAGTCCCAATGTGGA 2640  
Db 3080 GACGCTGTGTGTGGACCCCAAGAGAGAGACAAAGACTCTAAGGGATAGTCCCAATGTGGA 3139  
Qy 2641 GCGCGGCCATCTCGCGGAGACTTATTTATGGCTTGGCTGTGGAAACCAAGGATCTCTGCA 2700  
Db 3140 GCGCGGCCATCTCGCGGAGACTTATTTATGGCTTGGCTGTGGAAACCAAGGATCTCTGCA 3199  
Qy 2701 GCGCGGCCAGCTCGCGGAGAGAACTCTGGGCGCTCTCCCGACAGCATCTCTGCACTCGG 2760  
Db 3200 GCGCGGCCAGCTCGCGGAGAGAACTCTGGGCGCTCTCCCGACAGCATCTCTGCACTCGG 3259  
Qy 2761 AGAAACAACTCCAAAGATGAATCCCACTCTCTGTTTACGAGCGGTAGAGCTCGGCTGT 2820  
Db 3260 AGAAACAACTCCAAAGATGAATCCCACTCTCTGTTTACGAGCGGTAGAGCTCGGCTGT 3319  
Qy 2821 TTTTGTGCCAATTTGCTGACCAAAAGACTCTCTCTGGCGAAGAAATSCAAATGGCGG 2880  
Db 3320 TTTTGTGCCAATTTGCTGACCAAAAGACTCTCTCTGGCGAAGAAATSCAAATGGCGG 3379  
Qy 2881 ATATGTATGCAACGCGTGTGGCTCTACAGAGCTTCACTCGACTCCAGGCTTTTAA 2940  
Db 3380 ATATGTATGCAACGCGTGTGGCTCTACAGAGCTTCACTCGACTCCAGGCTTTTAA 3439  
Qy 2941 CATCATTTAAACAAACAAAGCTGAGCAGATTATTAGAGAGAGAACAAAGAGCGCTTAA 3000  
Db 3440 CATCATTTAAACAAACAAAGCTGAGCAGATTATTAGAGAGAGAACAAAGAGCGCTTAA 3499  
Qy 3001 CCCAGAGGCACTTCAGGCTGAGCAGTCAACAAACAGCAGAGGGGCGAGCAATGAGAGCA 3060  
Db 3500 CCCAGAGGCACTTCAGGCTGAGCAGTCAACAAACAGCAGAGGGGCGAGCAATGAGAGCA 3559  
Qy 3061 AGTCATTTGAGAGCCCTTACAGAGGAGGTACAGAGATCATCTAACTGAAAGTCAACAG 3120  
Db 3560 AGTCATTTGAGAGCCCTTACAGAGGAGGTACAGAGATCATCTAACTGAAAGTCAACAG 3619  
Qy 3121 AGAAATTCACACTCCCGAGCTTAAGTAAATACGAAGCCCAAGGCTTCACTGACTTAAAGCCA 3180  
Db 3620 AGAAATTCACACTCCCGAGCTTAAGTAAATACGAAGCCCAAGGCTTCACTGACTTAAAGCCA 3679  
Qy 3181 TTTTGTCTACAGCAGCAGTCTGTGTGAGCAACTCTGTGATATTCACAAAGATGCAACC 3240  
Db 3680 TTTTGTCTACAGCAGCAGTCTGTGTGAGCAACTCTGTGATATTCACAAAGATGCAACC 3739  
Qy 3241 TTTTGCATTTACAGTAAAGTCTCTCAGGAAAGTCTGAGATCCAGGAAATAGTTCATC 3300  
Db 3740 TTTTGCATTTACAGTAAAGTCTCTCAGGAAAGTCTGAGATCCAGGAAATAGTTCATC 3799  
Qy 3301 CGTATCTGAAAGGAAAGGAGTTCTCAGAGAGGAGCTCTCTATAGAAAGTACATGAGACC 3360  
Db 3800 CGTATCTGAAAGGAAAGGAGTTCTCAGAGAGGAGCTCTCTATAGAAAGTACATGAGACC 3859

```
Db      512  CAGAGGGGACGAAATGAGGAGCAAGTCAATGGAGCCCGTTAGAGAGGAGGTCAAGAGAT 453
Qy      3098 CATCTTAATGAAAGTCACAGAGAGAAATTCCTACTCCCAAGCCTAAGTAAATACGAAGCC 3157
Db      452  CATCTAACTGAAAGTCACAGAGAGAAATTCCTACTCCCAAGCCTAAGTAAATACGAAGCC 393
Qy      3158 CAGGGTTCATTGACTAAAGCCATTCTGCTCAGAGCCAGCTCTGCTCAGCCAACTCTG 3217
Db      392  CAGGGTTCATTGACTAAAGCCATTCTGCTCAGAGCCAGCTCTGCTCAGCCAACTCTG 333
Qy      3218 GATATTACAAAAGGATGCAACCTTTTGCACATTCAGATAAAAAGTCTCTCAGGAAGTACT 3277
Db      332  GATATTACAAAAGGATGCAACCTTTTGCACATTCAGATAAAAAGTCTCTCAGGAAGTACT 273
Qy      3278 GGAGATCCAGGAATAGTTCATCCGTATCTGAAGGGAAGAAAGTTCTGAGAGAGCCAGT 3337
Db      272  GGAGATCCAGGAATAGTTCATCCGTATCTGAAGGGAAGAAAGTTCCTGAGAGAGCCAGT 213
Qy      3338 CCTATAGAAAAGTACATGAGACCTGCGAAGACCCAAATTTATTCACCAAGGAGCCCT 3397
Db      212  CCTATAGAAAAGTACATGAGACCTGCGAAGACCCAAATTTATTCACCAAGGAGCCCT 153
Qy      3398 ATTGAAAAGTACCAGTACCACCTTTTGGACTTCCCTTTGCTACATATGACTTCCAGAGT 3457
Db      152  ATTGAAAAGTACCAGTACCACCTTTTGGACTTCCCTTTGCTACATATGACTTCCAGAGT 93
Qy      3458 GAAGCTGATTGGCTGGGTTCTGGAGTAAATATAAGCTCTCCCTGGGAATCCGCAC 3517
Db      92  GAAGCTGATTGGCTGGGTTCTGGAGTAAATATAAGCTCTCCCTGGGAATCCGCAC 33
Qy      3518 TACTTGAGTCACGCTGCGCTGCCCTACCAAAATCC 3549
Db      32  TACTTGAGTCACGCTGCGCTGCCCTACCAAAATCC 1
```

## RESULT 6

```
US-10-198-846-3994
; Sequence 3994, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3994
; LENGTH: 510
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-846-3994
```

```
Query Match
Best Local Similarity 10.1%; Score 453.4; DB 14; Length 510;
Matches 454; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2139 CATGTACCAATGATTTTATTACCAAGTGAAGAGAGATTTCCCGACACTACAGGA 2198
Db      56  CAGGTACCAATGATTTTATTACCAAGTGAAGAGAGATTTCCCGACACTACAGGA 115
Qy      2199 GAGCACACAGTCTGTACAAATGCGGTGAGTGCAGTTTACAGTGGCGGATACCTAGTCAC 2258
Db      116  GAGCACACAGTCTGTACAAATGCGGTGAGTGCAGTTTACAGTGGCGGATACCTAGTCAC 175
```

```
Qy      2259 TACTGGAGCACTTCAACACACTGTTTCACTGCCAGGAACAGGACATCACTACAGCCACGGCG 2318
Db      176  TACTGGAGCACTTCAACACACTGTTTCACTGCCAGGAACAGGACATCACTACAGCCACGGCG 235
Qy      2319 AAGAGGACGGTTCATGCCATATCCACCATCAAGAGAGGAGCCCAAAATTTGACTTCAGGGTCT 2378
Db      236  AAGAGGACGGTTCATGCCATATCCACCATCAAGAGAGGAGCCCAAAATTTGACTTCAGGGTCT 295
Qy      2379 ACAATCTGCTTAACCTCCAGACTCTTAAATGGAGAGCCAGTTTCTGAAGTGTGTGAAGA 2438
Db      296  ACAATCTGCTTAACCTCCAGACTCTTAAATGGAGAGCCAGTTTCTGAAGTGTGTGAAGA 355
Qy      2439 GAGAGAGCTGGAGAGAGAGACCGGCTCAAGAGAGAAAGTTTGGACCSAGAGTTCAGTG 2498
Db      356  GAGAGAGCTGGAGAGAGAGACCGGCTCAAGAGAGAAAGTTTGGACCSAGAGTTCAGTG 415
Qy      2499 ATGACCTTTCGCAATGTGACTTTGGAGAGGGGACAGACATCTCGCGGGGAGTCCGTACACA 2558
Db      416  ATGACCTTTCGCAATGTGACTTTGGAGAGGGGACAGACATCTCGCGGGGAGTCCGTACACA 475
Qy      2559 CCCAAGCAAGCTGGGCTGCTGACGGCTGTGTCT 2593
Db      476  CCCAAGCAAGCTGGGCTGCTGACGGCTGTGTCT 510
```

## RESULT 7

```
US-10-198-846-8958
; Sequence 8958, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8958
; LENGTH: 523
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2,3,5
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-8958
```

```
Query Match
Best Local Similarity 8.9%; Score 398.8; DB 14; Length 520;
Matches 400; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1418 GGTACAGAGGCCACAGTTACTACTGTTGTAATTTTGTAGTTTACGTGTGAGTCACT 1477
Db      111  GGTACAGAGGCCACAGTTACTACTGTTGTAATTTTGTAGTTTACGTGTGAGTCACT 170
Qy      1478 AGCTCACTTAACCTGCTAGACATATGCGAAGCAGCAGCAGCAGTCCAGTCCAGCGCG 1537
Db      171  AGCTCACTTAACCTGCTAGACATATGCGAAGCAGCAGCAGTCCAGTCCAGCGCG 230
Qy      1538 CTTAATCCAGAGTTAAATGATAAGCTTTCCAGGGGCTCTGTCACTTAATCAGAAATGACTA 1597
Db      231  CTTAATCCAGAGTTAAATGATAAGCTTTCCAGGGGCTCTGTCACTTAATCAGAAATGACTA 290
Qy      1598 GCCAAAAGTTTCAGAGGAGAGACAAATGACCAAGACAGACAAAGAGCTCAGTGGGGCTAAA 1657
Db      291  GCCAAAAGTTTCAGAGGAGAGACAAATGACCAAGACAGACAAAGAGCTCAGTGGGGCTAAA 350
```



QY 1658 AAGAGGACTTCTCCAGCAGGAGCGGAGGATATATGTTAACGAGCTATATTTTCAG 1717  
 |||||  
 DB 351 AAGAGGACTTCTTAGCAGGAGGAGCGGAGGATATATGTTAACGAGCTATATTTTCAG 410  
 |||||  
 QY 1718 TTCTGTGACTTCCGATATATCCAAAGCCATGGCCCTGATGTAATTTAGTGGGGCACTT 1777  
 |||||  
 DB 411 TTCTGTGACTTCCGATATATCCAAAGCCATGGCCCTGATGTAATTTAGTGGGGCACTT 470  
 |||||  
 QY 1778 CTCGGCTATATCAACAGCTCCATCAATCAATCAAGTGATCC 1819  
 |||||  
 DB 471 CTCGGCTATATCAACAGCTCCATCAATCAATCAAGTGATCC 512  
 |||||

## RESULT 8

US-10-198-846-3330/C  
 ; Sequence 3330, Application US/10198846  
 ; Publication No. US2003009974A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lillie, James  
 ; APPLICANT: Xu, Yongyao  
 ; APPLICANT: Wang, Youzhen  
 ; APPLICANT: Steinhmann, Kathleen  
 ; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
 ; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
 ; TITLE OF INVENTION: THERAPY OF BREAST CANCER  
 ; FILE REFERENCE: MRI-049  
 ; CURRENT APPLICATION NUMBER: US/10/198,846  
 ; CURRENT FILING DATE: 2002-07-18  
 ; PRIOR APPLICATION NUMBER: 60/306,220  
 ; PRIOR FILING DATE: 2001-07-18  
 ; NUMBER OF SEQ ID NOS: 14084  
 ; SOFTWARE: Fast-SEQ for Windows Version 4.0  
 ; SEQ ID NO 3330  
 ; LENGTH: 836  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: 3, 23, 27, 28, 622, 624, 625, 628, 629, 630, 631, 633, 634,  
 ; LOCATION: 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646,  
 ; LOCATION: 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658,  
 ; LOCATION: 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669  
 ; OTHER INFORMATION: n = A,T,C or G  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681,  
 ; LOCATION: 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693,  
 ; LOCATION: 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705,  
 ; LOCATION: 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716  
 ; OTHER INFORMATION: n = A,T,C or G  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728,  
 ; LOCATION: 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740,  
 ; LOCATION: 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752,  
 ; LOCATION: 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763  
 ; OTHER INFORMATION: n = A,T,C or G  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775,  
 ; LOCATION: 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787,  
 ; LOCATION: 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799,  
 ; LOCATION: 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810  
 ; OTHER INFORMATION: n = A,T,C or G  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822,  
 ; LOCATION: 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834,  
 ; LOCATION: 835  
 ; OTHER INFORMATION: n = A,T,C or G

US-10-198-846-3330

Query Match 8.6%; Score 389.2; DB 14; Length 836;  
 Best Local Similarity 96.5%; Pred. No. 1.7e-101; Indels 2; Gaps 1;  
 Matches 409; Conservative 0; Mismatches 13;  
 QY 1396 CTTGGATTCCTCTAGACAAATGTTACAGAGCCACCACTTACTACTGTGTAAATTTTG 1455  
 |||||  
 DB 486 CCACTAGTTCTAGAGCGCGGAGGTACAGAGGCCACCACTTACTACTGTGTAAATTTTG 427  
 |||||  
 QY 1456 TAGTTTCAGCTGTAGTCTATCTAGCTCATTAACTGCTAGACATTATGCGAGCAGCA 1515  
 |||||  
 DB 426 TAGTTTCAGCTGTAGTCTATCTAGCTCATTAACTGCTAGACATTATGCGAGCAGCA 367  
 |||||  
 QY 1516 CCGAGCAGTGCAGTTCAGGGGGCTTAATCCAGAGTTAAATGATTAAGCTTCCAGGGGCTC 1575  
 |||||  
 DB 366 CCGAGCAGTGCAGTTCAGGGGGCTTAATCCAGAGTTAAATGATTAAGCTTCCAGGGGCTC 307  
 |||||  
 QY 1575 TGTCAATTAATCAGATATGATCTAGCCAAAAGTTTCAAGAGGAGAGACAATGACCAAGACAGA 1635  
 |||||  
 DB 306 TGTCAATTAATCAGATATGATCTAGCCAAAAGTTTCAAGAGGAGAGACAATGACCAAGACAGA 247  
 |||||  
 QY 1636 CAGAGCTCGAGTGGGGCTAATAAGAGAGAGCTCTCCAGGAGGAGGAGGAGGATATAT 1695  
 |||||  
 DB 246 CAGAGCTCGAGTGGGGCTAATAAGAGAGAGCTCTCCAGGAGGAGGAGGAGGATATAT 187  
 |||||  
 QY 1696 GGTAAAGAGCTATAATTTGTCAGTTCTGTGACTTCGGATATTCCTAAAGCCATGSCCTTGA 1755  
 |||||  
 DB 196 GG--ACGAGCTATAATTTGTCAGTTCTGTGACTTCGGATATTCCTAAAGCCATGSCCTTGA 129  
 |||||  
 QY 1756 TGTAAATGTAGTGGGCGCACTTCTCGTCATTATCAACAGCTCCATTAACATTCAAGTG 1815  
 |||||  
 DB 128 TGTAAATGTAGTGGGCGCACTTCTCGTCATTATCAACAGCTCCATTAACATTCAAGTG 69  
 |||||  
 QY 1816 TACC 1819  
 |||||  
 DB 68 TACC 65

## RESULT 9

US-10-198-846-5878  
 ; Sequence 5878, Application US/10198846  
 ; Publication No. US2003009974A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lillie, James  
 ; APPLICANT: Xu, Yongyao  
 ; APPLICANT: Wang, Youzhen  
 ; APPLICANT: Steinhmann, Kathleen  
 ; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
 ; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
 ; TITLE OF INVENTION: THERAPY OF BREAST CANCER  
 ; FILE REFERENCE: MRI-049  
 ; CURRENT APPLICATION NUMBER: US/10/198,846  
 ; CURRENT FILING DATE: 2002-07-18  
 ; PRIOR APPLICATION NUMBER: 60/306,220  
 ; PRIOR FILING DATE: 2001-07-18  
 ; NUMBER OF SEQ ID NOS: 14084  
 ; SOFTWARE: Fast-SEQ for Windows Version 4.0  
 ; SEQ ID NO 5878  
 ; LENGTH: 427  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: 104, 128, 132, 145, 158, 167, 173, 177, 207, 210, 250, 254,  
 ; LOCATION: 296, 308, 324, 389, 394, 398, 407  
 ; OTHER INFORMATION: n = A,T,C or G  
 ; OTHER INFORMATION: n = A,T,C or G

US-10-198-846-5878

Query Match 7.2%; Score 324.4; DB 14; Length 427;  
 Best Local Similarity 91.7%; Pred. No. 6.1e-83;  
 Matches 164; Conservative 0; Mismatches 30; Indels 3; Gaps 3;  
 QY 1416 GGTACAGAGCCACCACTTACTGTTAAATTTGTAGTTCAAGTGTGAGTCATCT 1477  
 |||||



```
Db 27 GGTACAGAGGCC-CCAGTTACTGCTGTAATTTTGTAGTTTCAGCTGAGTCATCT 85
QY 1478 AGCTACTTAAGTCTGAGAAATATGCGAAGCAGCAGGAGCAGTCCAGTCAGGGGC 1537
Db 86 AGCTACTTAAGTCTGAGAAATATGCGAAGCAGCAGGAGCAGTCCAGTCCAGGGGC 145
QY 1538 CTTAATCCAGCTTAATGATGAGCTTTCCAGGGGCTGTGTCATTAATC-AGATGATCT 1596
Db 146 CTTAATCCAGCTTAATGATGAGCTTTCCAGGGGCTGTGTCATTAATCAAGAATGATCT 205
QY 1597 AGCCAAAGTTTCAGAGGAGACATGATGACCAAGACAGACAGAGCTCGAGTGGGGCTAA 1656
Db 206 ANCCNAAAGTTTCAGAGGAGACATGATGACCAAGACAGACAGAGCTCNAGTGGGGCTAA 265
QY 1657 AAGAGAGGACTTTCAGCAAGAGGAG-CCGAGGATAATATGATACGAGCTATTAATGTC 1715
Db 266 AAGAGAGGACTTTCAGCAAGAGGAGCCCGGAGATAATATGAGACAGCTATTAATGNC 325
QY 1716 AGTTCTGTGACTTCCGATATCCAAAGCCATGGCCCTGATTAATTTAGTGGGGCCAC 1775
Db 326 AGCCCTGTGACTTCCGATATCCAAAGCCATGGCCCTGATTAATTTAGTGGGGCCAC 385
QY 1776 TTCTCGCTGATTAATCAACAGCTCCATAACATTACAA 1812
Db 386 TTCNCCGTNATTCACAGCTGTCATTACATTACAA 422

RESULT 10
US-09-918-995-10414
; Sequence 10414, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10414
; LENGTH: 476
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(476)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-10414

Query Match 7.2%; Score 322.8; DB 11; Length 476;
Best Local Similarity 96.2%; Pred. No. 1.9e-82;
Matches 327; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4162 AGTGAGTAATAGTCAATGAGAAAGATGGAAGATGGTAAACATCACTTTTAAACC 4221
Db 24 AGNGCANTCCNNGNCNGAGAAAGATGGAAGATGGTAAACATCACTTTTAAACC 83
QY 4222 TGTTAAGTCAAAACCATCTTGGCTAATATGCTAGGGGAAATATCCATAGAGATATCA 4281
Db 84 TGTTAAGTCAAAACCATCTTGGTAAATATGCTAGGGGAAATATCCATAGAGATATCA 143
QY 4282 CCAGCTAGAAATTAATATTTTAAAGAAAGAGACCAAAATGCTAGAAATTTGAAAGG 4341
Db 144 CCAGCTAGAAATTAATATTTTAAAGAAAGAGACCAAAATGCTAGAAATTTGAAAGG 203
QY 4342 GTTTTACATATTTATATTAATGAGCAGTCTGGCTGGCCATGGACATTTGTTCAAAA 4401
Db 204 GTTTTACATATTTATATTAATGAGCAGTCTGGCTGGCCATTTGGACCATTTGTTCAAAA 263
QY 4402 CCCATAAATTTGCTTAAATTTTATAATGATCATGAACCCCTAGGCAGAGGAGGAAT 4461
```

```
Db 264 CCATATAATTTGCTTAAATTTATATGATCATGAACCCCTAGGCAGAGGAGGAAT 323
QY 4462 TGAAGTCCAGGGCAATGAAAGAAAATGGCCCTCTCA 4501
Db 324 TGAAGTCCAGGGCAATGAAAGAAAATGGCCCTCTCA 363

RESULT 11
US-09-778-320-110
; Sequence 110, Application US/09778320
; Patent No. US20030034052A1
; GENERAL INFORMATION:
; APPLICANT: Billion, David C.
; APPLICANT: Day, Craig H.
; APPLICANT: Jiang, Yugu
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Mitcham, Jennifer
; APPLICANT: Wang, TongTong
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 21021.491CS
; CURRENT APPLICATION NUMBER: US/09/778,320
; CURRENT FILING DATE: 2001-02-06
; NUMBER OF SEQ ID NOS: 301
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 110
; LENGTH: 325
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(305)
; OTHER INFORMATION: n = A,T,C or G
US-09-778-320-110

Query Match 6.1%; Score 275.4; DB 9; Length 305;
Best Local Similarity 94.1%; Pred. No. 7.6e-69;
Matches 287; Conservative 0; Mismatches 17; Indels 1; Gaps 1;

QY 3439 ACATATGACTTCCAGAGTGAAGCTGATGGCTGGGTCTGGAGTAATATAAGCTCTC 3498
Db 3439 ACATATGACTTCCAGAGTGAAGCTGATGGCTGGGTCTGGAGTAATATAAGCTCTC 60
QY 3499 GCTTCCTGGGAATCCGCACTACTTGAGTCACTGGCTGGCTGACCAATCCCTGCCAAA 3558
Db 61 GCTTCCTGGGAATCCGCACTACTTGAGTCACTGGCTGGCTGACCAATCCCTGCCAAA 120
QY 3559 CTATGTCCTTATCCACCTT-CAATCTGGCTCTCATTTTCAGCTCTGGATCAGACA 3617
Db 121 CTATGTCCTTATCCACCTTNNAACTGNSCTCTCTCACTTNTCAGCTCTGGATCAGACA 180
QY 3618 ATGACATTCCTCTAGATTTGGGATCAAGCATTCAGACCTGGGCAACTGCAAAACGGT 3677
Db 181 ATGACATTCCTCTAGATTTGGGATCAAGCATTCAGACCTGGGCAACTGCAAAACGGT 240
QY 3678 CTCCAGGAGAAAACGAGAGCCACCAATGTAATAAATGAAGTCCCTTGAATGAG 3737
Db 241 CTTCNAGGAGAAAACGAGAGCCACCAATGTAATAAATGAAGTCCCTTGAATGAG 300
QY 3738 TAAAA 3742
Db 301 TAAAA 305

RESULT 12
US-09-510-689-110
; Sequence 110, Application US/09910689
; Patent No. US20020081609A1
; GENERAL INFORMATION:
; APPLICANT: Billion, David C.
; APPLICANT: Day, Craig H.
```

```
APPLICANT: Jiang, Yugu
APPLICANT: Houghton, Raymond L.
APPLICANT: Mitcham, Jennifer
APPLICANT: Wang, Tongtong
APPLICANT: McNeill, Patricia D.
APPLICANT: Harlocker, Susan L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.491C6
CURRENT APPLICATION NUMBER: US/09/910,689
CURRENT FILING DATE: 2001-07-20
NUMBER OF SEQ ID NOS: 307
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 110
LENGTH: 305
TYPE: DNA
ORGANISM: Homo sapiens
NAME/KEY: misc feature
LOCATION: 12, 13, 16, 110, 142, 143, 150, 161, 192, 198, 217, 223,
LOCATION: 244, 263, 274, 285, 287
OTHER INFORMATION: n = A,T,C or G
US-09-910-689-110

Query Match      6.1%  Score 275.4; DB 9; Length 305;
Best Local Similarity 94.1%  Pred. No. 7.6e-69;
Matches 287; Conservative 0; Mismatches 17; Indels 1; Gaps 1;

QY 3439 ACATAATGACTTCCAGAGTGAAGCTGATGGCTGGCTCTGAGTAATATTAAGCTCTC 3498
Db 1 ACATAATGACTTCCAGAGTGAAGCTGATGGCTGGCTCTGAGTAATATTAAGCTCTC 60

QY 3499 CGTTCCTGGGAATCCGCACTACTTGAGTCAGCTGCGCTACCAAACTCCTTGCCAAA 3558
Db 61 CGTTCCTGGGAATCCGCACTACTTGAGTCAGCTGCGCTACCAAACTCCTTGCCAAA 120

QY 3559 CTATGTGCCCTTATCCACCTT-CAATCTGCTCTCTCATTTTTCAGCTGTGGATCAGACA 3617
Db 121 CTATGTGCCCTTATCCACCTTNNAACTGNCCTCTCATTTTTCAGCTGTGGATCAGACA 180

QY 3618 ATGACATTCCTCTAGATTGGCGATCAAGCATTCACAGCTGGCGCAACTGCAAAACGGTG 3677
Db 181 ATGACATTCCTCTAGATTGGCGATCAAGCATTCACAGCTGGCGCAACTGCAAAACGGTG 240

QY 3678 CCTCAGGAGAAAACGAGGACCACCAATGTAAATAATGAAGTCCCTTGAATGTAG 3737
Db 241 CCTCAGGAGAAAACGAGGACCACCAATGTAAATAATGAAGTCCCTTGAATGTAG 300

QY 3738 TAAAA 3742
Db 301 TAAAA 305

RESULT 13
US-10-010-742-110
Sequence 110, Application US/10010742
Publication No. US2002014672A1
GENERAL INFORMATION:
APPLICANT: Dillon, Davin C.
APPLICANT: Day, Craig H.
APPLICANT: Jiang, Yugu
APPLICANT: Houghton, Raymond L.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Wang, Tongtong
APPLICANT: McNeill, Patricia D.
APPLICANT: Harlocker, Susan L.
APPLICANT: Bennington, Angela Ann
APPLICANT: Zehentner, Barbara
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.491C7
```

```
CURRENT APPLICATION NUMBER: US/10/010,742
CURRENT FILING DATE: 2001-11-30
NUMBER OF SEQ ID NOS: 307
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 110
LENGTH: 305
TYPE: DNA
ORGANISM: Homo sapiens
NAME/KEY: misc feature
LOCATION: 12, 13, 16, 110, 142, 143, 150, 161, 192, 198, 217, 223,
LOCATION: 244, 263, 274, 285, 287
OTHER INFORMATION: n = A,T,C or G
US-10-010-742-110

Query Match      6.1%  Score 275.4; DB 13; Length 305;
Best Local Similarity 94.1%  Pred. No. 7.6e-69;
Matches 287; Conservative 0; Mismatches 17; Indels 1; Gaps 1;

QY 3439 ACATAATGACTTCCAGAGTGAAGCTGATGGCTGGCTCTGAGTAATATTAAGCTCTC 3498
Db 1 ACATAATGACTTCCAGAGTGAAGCTGATGGCTGGCTCTGAGTAATATTAAGCTCTC 60

QY 3499 CGTTCCTGGGAATCCGCACTACTTGAGTCAGCTGCGCTACCAAACTCCTTGCCAAA 3558
Db 61 CGTTCCTGGGAATCCGCACTACTTGAGTCAGCTGCGCTACCAAACTCCTTGCCAAA 120

QY 3559 CTATGTGCCCTTATCCACCTT-CAATCTGCTCTCTCATTTTTCAGCTGTGGATCAGACA 3617
Db 121 CTATGTGCCCTTATCCACCTTNNAACTGNCCTCTCATTTTTCAGCTGTGGATCAGACA 180

QY 3618 ATGACATTCCTCTAGATTGGCGATCAAGCATTCACAGCTGGCGCAACTGCAAAACGGTG 3677
Db 181 ATGACATTCCTCTAGATTGGCGATCAAGCATTCACAGCTGGCGCAACTGCAAAACGGTG 240

QY 3678 CCTCAGGAGAAAACGAGGACCACCAATGTAAATAATGAAGTCCCTTGAATGTAG 3737
Db 241 CCTCAGGAGAAAACGAGGACCACCAATGTAAATAATGAAGTCCCTTGAATGTAG 300

QY 3738 TAAAA 3742
Db 301 TAAAA 305

RESULT 14
US-10-198-846-7424
Sequence 7424, Application US/10198845
Publication No. US2003009974A1
GENERAL INFORMATION:
APPLICANT: Lillie, James
APPLICANT: Xu, Yongyao
APPLICANT: Wang, Youzhen
APPLICANT: Steinhart, Kathleen
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
FILE REFERENCE: MFI-049
TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
FILE REFERENCE: MFI-049
CURRENT APPLICATION NUMBER: US/10/198,846
CURRENT FILING DATE: 2002-07-18
PRIOR APPLICATION NUMBER: 60/306,220
PRIOR FILING DATE: 2001-07-18
NUMBER OF SEQ ID NOS: 14084
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 7424
LENGTH: 810
TYPE: DNA
ORGANISM: Homo sapiens
NAME/KEY: misc feature
LOCATION: 2, 3, 4, 5, 6, 13, 434, 435, 436, 437, 438, 439, 440, 441,
LOCATION: 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453,
LOCATION: 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465,
LOCATION: 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476
```

```
/ OTHER INFORMATION: n = A,T,C or G
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488,
/ LOCATION: 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500,
/ LOCATION: 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512,
/ LOCATION: 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523
/ OTHER INFORMATION: n = A,T,C or G
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535,
/ LOCATION: 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547,
/ LOCATION: 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559,
/ LOCATION: 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570
/ OTHER INFORMATION: n = A,T,C or G
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582,
/ LOCATION: 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594,
/ LOCATION: 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606,
/ LOCATION: 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617
/ OTHER INFORMATION: n = A,T,C or G
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629,
/ LOCATION: 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641,
/ LOCATION: 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653,
/ LOCATION: 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664
/ OTHER INFORMATION: n = A,T,C or G
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676,
/ LOCATION: 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688,
/ LOCATION: 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700,
/ LOCATION: 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711-810
/ OTHER INFORMATION: n = A,T,C or G
US-10-198-846-7424

Query Match 4.7% Score 210; DB 14; Length 810;
Best Local Similarity 100.0%; Pred.No. 1.5e-49;
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1815 GTACATTAAACACTGCCATTCTCTCCAGAGACTTGTCCAGCCAGACCCAGAAAGCACCTTG 1874
DB 113 GTACATTAAACACTGTCCATTCTCTCCAGAGACTTGTCCAGCCAGACCCAGAAAGCACCTTG 172
QY 1875 GAGAAATTAATCTATCCGTTTCTGTAGAAAAGTAATTTGTCCACTGTGCACCTTTGC 1934
DB 173 GAGAAATTAATCTATCCGTTTCTGTAGAAAAGTAATTTGTCCACTGTGCACCTTTGC 232
QY 1935 TTCTGCACCTTGTCTCTCTGGGCGGCTGGAAGCTCGCAGTCAAAACATCAGTGCACCTCAGT 1994
DB 233 TTCTGCACCTTGTCTCTCTGGGCGGCTGGAAGCTCGCAGTCAAAACATCAGTGCACCTCAGT 252
QY 1995 GTTCATTCACACCCCTGCAGTAGATGTAC 2024
DB 293 GTTCATTCACACCCCTGCAGTAGATGTAC 322

RESULT 15
US-10-198-846-9634
; Sequence 9634, Application US/10198846
; Publication No. US20030099741
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846

/ CURRENT FILING DATE: 2002-07-18
/ PRIOR APPLICATION NUMBER: 60/306,223
/ PRIOR FILING DATE: 2001-07-15
/ NUMBER OF SEQ ID NOS: 14284
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 9634
/ LENGTH: 594
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 1, 5, 27, 32, 35, 38, 45, 46, 69, 319, 355, 385, 395, 401,
/ LOCATION: 414, 416, 420, 424, 425, 427, 428, 429, 430, 431, 432, 433,
/ LOCATION: 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445,
/ LOCATION: 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456
/ OTHER INFORMATION: n = A,T,C or G
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 457, 459, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468,
/ LOCATION: 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480,
/ LOCATION: 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492,
/ LOCATION: 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503
/ OTHER INFORMATION: n = A,T,C or G
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515,
/ LOCATION: 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527,
/ LOCATION: 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539,
/ LOCATION: 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550
/ OTHER INFORMATION: n = A,T,C or G
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562,
/ LOCATION: 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574,
/ LOCATION: 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586,
/ LOCATION: 587, 588, 589, 590, 591, 592, 593, 594
/ OTHER INFORMATION: n = A,T,C or G
US-10-198-846-9634

Query Match 3.0%; Score 133.4; DB 14; Length 594;
Best Local Similarity 99.3%; Pred.No. 1.9e-27;
Matches 134; Conservative 0; Mismatches 1; Indels 3; Gaps 0;

QY 4367 GTACTGGACTGCCATTGGACCATTTGTTCCAAAACCCATTAATTTGTCCTAAATTTAT 4426
DB 142 GTACTGGACCGCCATTGGACCATTTGTTCCAAAACCCATTAATTTGTCCTAAATTTAT 201
QY 4427 AATGATTCATGAAACCCCTAGGCGAGGAGGAGAAATTTGAAGTTCAGGCGCAATGAAGAAA 4486
DB 202 AATGATTCATGAAACCCCTAGGCGAGGAGGAGAAATTTGAAGTTCAGGCGCAATGAAGAAA 261
QY 4487 AATGGCGCCCTCTCA 4501
DB 262 AATGGCGCCCTCTCA 276

Search completed: October 29, 2003, 16:03:38
Job time : 1:30 secs
```